



Universidad  
Carlos III de Madrid

Departamento de Informática

PROYECTO FIN DE CARRERA DE  
INGENIERÍA TÉCNICA DE GESTIÓN

# **HERRAMIENTA DE SOPORTE A LA DECISIÓN: INFLUENCIA DE LAS NOTICIAS FINANCIERAS EN BOLSA**

Autor: Ismael Rodrigo Huete

Tutor: Dolores Cuadra Fernández

Leganés, Octubre de 2015



Título: HERRAMIENTA DE SOPORTE A LA DECISIÓN: INFLUENCIA DE LAS NOTICIAS FINANCIERAS EN BOLSA

Autor: Ismael Rodrigo Huete

Director: Dolores Cuadra Fernández

## EL TRIBUNAL

Presidente: Jorge Morato

Vocal: Elena Castro

Secretario: Harith Aljumaily

Realizado el acto de defensa y lectura del Proyecto Fin de Carrera el día 26 de Octubre de 2015 en Leganés, en la Escuela Politécnica Superior de la Universidad Carlos III de Madrid, acuerda otorgarle la CALIFICACIÓN de

VOCAL

SECRETARIO

PRESIDENTE



# Agradecimientos

Agradezco a mi familia y amigos.

# Resumen

Actualmente se procesa un gran volumen de información. La información siempre ha aportado un valor añadido y es clave en la toma de decisiones.

Basandonos en dos aspectos primordiales de la información, la fidelidad de la información obtenida y su correcta interpretación, podemos llegar a prever con cierta aproximación cualquier fenómeno. Por ejemplo, podemos saber con cierta exactitud al lanzar una moneda al aire, si saldrá cara o cruz. Esto es posible si disponemos de toda la información que influya en este fenómeno (potencia de lanzamiento, dimensiones y peso de la moneda, parámetros ambientales, gravedad, entre otros).

Las empresas y grandes inversores siempre han intentado encontrar un patrón o síntomas que puedan ‘avisar’ o prever el comportamiento de la economía y los mercados de valores. Para ello se realizan estudios sobre el comportamiento histórico que puedan dar una cierta aproximación. Hoy en día con la gran cantidad de información que disponemos, este proceso es más complejo, pero también más fiable.

La casualidad no se puede considerar un factor clave en operaciones de grandes cantidades de dinero. Este proyecto busca conocer en qué medida impacta la gran cantidad de información que disponemos en los mercados financieros y ayudar a pequeños usuarios a poder realizar inversiones basándose en recomendaciones automáticas.

**Palabras clave:** Análisis de sentimiento, mercado de valores, noticias

# Abstract

Currently a large volume of information is processed. Information has always brought added value and is key in decision-making.

Basing on two key aspects of the information (the accuracy of the information obtained and its proper interpretation) we can foresee any phenomenon with some approximation. For example, we may know with some accuracy by throwing a coin, if it will be heads or tails. This is possible if we have all the information to influence this phenomenon (launch power, size and weight of the coin, environmental parameters, gravity, etc.).

Companies and large investors have always tried to find a pattern or symptoms that might 'warn' or predict the behavior of the economy and the stock markets. In this sense, researches on the historical behavior that may give some approximation are performed. Today with the wealth of information we have, this process is more complex, but also more reliable.

The accident can not be considered a key factor in operations of large amounts of money. This project seeks to understand to what extent the wealth of information that we have impacts on financial markets and help small users to make investments based on automatic recommendations.

**Keywords:** Sentimental Analysis, stock market, news

# Índice general

<b>1. INTRODUCCIÓN Y OBJETIVOS .....</b>	<b>11</b>
1.1 Introducción .....	11
1.2 Motivación .....	11
1.3 Objetivos del proyecto .....	12
1.3.1 <i>Objetivo principal</i> .....	12
1.3.2 <i>Objetivos secundarios</i> .....	13
1.4 Estructura del documento .....	13
<b>2. ESTADO DEL ARTE.....</b>	<b>15</b>
2.1 Introducción .....	15
2.2 Proyectos similares.....	16
2.2.1 <i>Estudio de la Universidad de Warwick</i> .....	16
2.2.2 <i>Estudio de las Universidades de Londres y Boston</i> .....	17
2.3 Fuentes de datos .....	19
2.3.1 <i>Datos Bolsa de Madrid 2014</i> .....	19
2.3.2 <i>Noticias de 2014</i> .....	20
2.4 Análisis de sentimiento .....	21
2.4.1 <i>Semantria</i> .....	22
2.4.2 <i>Alchemy API</i> .....	24
2.4.3 <i>Lymbix</i> .....	25
2.4.4 <i>Repustate</i> .....	26
2.4.5 <i>MeaningCloud Add-in para Excel</i> .....	31
2.4.6 <i>Otras</i> .....	39
2.5 Análisis y visualización de datos .....	39
2.5.1 <i>Análisis</i> .....	39
2.5.2 <i>Análisis de noticias</i> .....	39
2.6 Metodología SCRUM .....	39
<b>3. ANÁLISIS, DISEÑO E IMPLEMENTACIÓN DEL SISTEMA .....</b>	<b>41</b>
3.1 Análisis de requisitos de usuario .....	41
3.1.1 <i>Requisitos funcionales</i> .....	42
3.1.2 <i>Requisitos no funcionales</i> .....	42
3.1 Diseño e implementación .....	42
<b>4. PRUEBAS.....</b>	<b>45</b>
4.1 Pruebas funcionales.....	45
4.2 Pruebas no funcionales.....	51
<b>5. PLANIFICACIÓN.....</b>	<b>52</b>
5.1 Ciclo de vida iterativo .....	53
5.2 Planificación inicial.....	55
5.3 Ejecución final del proyecto.....	59
<b>6. PRESUPUESTO .....</b>	<b>64</b>
6.1 Presupuesto inicial.....	65
6.1.1 <i>Costes de Personal</i> .....	65
6.1.2 <i>Costes de Equipamiento</i> .....	66
6.1.3 <i>Coste Total</i> .....	66
6.2 Coste real.....	67
6.2.1 <i>Costes de Personal</i> .....	67
6.2.2 <i>Costes de Equipamiento</i> .....	67
6.2.3 <i>Coste Total</i> .....	67
<b>7. CONCLUSIONES .....</b>	<b>69</b>
<b>8. TRABAJOS FUTUROS.....</b>	<b>70</b>



<b>9. GLOSARIO DE ACRÓNIMOS .....</b>	<b>71</b>
<b>10. BIBLIOGRAFÍA.....</b>	<b>72</b>
<b>11. ANEXO I. CÁLCULO SALARIO PROFESIONALES.....</b>	<b>73</b>

# Índice de Ilustraciones

Ilustración 1: Gráfico que compara la cotización del Ibex 35 con las búsquedas en Google (Fuente: Blog Bankinter).....	18
Ilustración 2: Ejemplo de un registro del cierre de Bolsa .....	20
Ilustración 3: Ejemplo de un registro de la tabla Noticia.....	21
Ilustración 4: Página web principal de Semantria for Excel <a href="https://semantria.com/excel">https://semantria.com/excel</a> .....	23
Ilustración 5: Página web principal de Semantria API <a href="https://semantria.com/API">https://semantria.com/API</a> .....	23
Ilustración 6: Página web principal de AlchemyAPI <a href="http://www.alchemyapi.com/">http://www.alchemyapi.com/</a> .....	24
Ilustración 7: Página web AlchemyAPI. Demo web.....	25
Ilustración 8: Página web principal de Lymbix. <a href="http://www.lymbix.com/">http://www.lymbix.com/</a> .....	25
Ilustración 9: Página web Lymbyx. Sign Up free trial.....	26
Ilustración 10: Página web Lymbyx. Live Demo .....	26
Ilustración 11: Página web principal de Repustate. <a href="http://www.repustate.com/">http://www.repustate.com/</a> .....	29
Ilustración 12: Página web Repustate. Demo on-line .....	30
Ilustración 13: Página web principal de MeaningCloud. <a href="https://www.meaningcloud.com/es/">https://www.meaningcloud.com/es/</a> .....	31
Ilustración 14: Página web de registro en MeaningCloud .....	32
Ilustración 15: Contador de request de MeaningCloud.....	32
Ilustración 16: Ribbon MeaningCloud Add-in Excel .....	33
Ilustración 17: Input Sentiment Analysis .....	35
Ilustración 18: Advanced Settings. Sentiment Analysis .....	36
Ilustración 19: Polarity Tags .....	37
Ilustración 20: Ejemplo Global Sentiment Analysis.....	38
Ilustración 21: Ejemplo Topic Sentiment Analysis.....	38
Ilustración 22: Pantalla inicial “Analizador Bolsa” .....	43
Ilustración 23: Gráfico de la aplicación “Analizador Bolsa” .....	44
Ilustración 24: Prueba 1.1.....	46
Ilustración 25: Prueba 1.2.....	46
Ilustración 26: Prueba 1.3.....	47
Ilustración 27: Prueba 1.4.....	47
Ilustración 28: Prueba 2.1.....	48
Ilustración 29: Prueba 2.2.....	48
Ilustración 30: Prueba 2.3.....	49
Ilustración 31: Prueba 4.1.....	50
Ilustración 32: Prueba 4.2.....	50
Ilustración 33: Esque del Ciclo de Vida Iterativo .....	54
Ilustración 34: Costes de equipamiento .....	57
Ilustración 35: Diagrama de Gantt. Planificación estimada .....	58
Ilustración 36: Costes de equipamiento .....	62
Ilustración 37: Diagrama de Gantt. Planificación final.....	63
Ilustración 38: Costes de Personal .....	66
Ilustración 39: Costes de equipamiento .....	66
Ilustración 40: Costes de equipamiento .....	67
Ilustración 41: Cálculo salario Ingeniero Técnico Informático Jr.....	73
Ilustración 42: Cálculo salario Ingeniero Informático Sr.....	74

# Capítulo 1

## Introducción y objetivos

### 1.1 Introducción

A lo largo de este documento se va a realizar un análisis del estado actual de la cuestión, analizando el comportamiento del mercado de valores en función de las noticias. Así mismo, se estudiará la forma de analizar los datos extraídos para convertirlos en información útil y visualmente atractiva a la hora de tomar decisiones.

Habrà una descripción de los principales métodos de obtención y tratamiento de esa información y, en concreto, la explicación de cómo se ha creado una herramienta para visualizar información a partir de una base de datos y la transforma de tal manera que se pueda obtener una conclusión (positivo/negativo) de estos.

Al realizar una clasificación de todas las noticias relacionadas con empresas del Ibex35, se podrá obtener un balance general basado en cifras exactas de cómo afectan las noticias al movimiento de los mercados. Se realizará una herramienta de visualización adaptada a un usuario no experto.

### 1.2 Motivación

La alta capacidad de procesamiento de información disponible hoy en día hace posible plantearnos predecir el futuro en cualquier ámbito, dejando atrás las “casualidades”. Basándome en esta inquietud, el principal motivo que me ha llevado a realizar un trabajo relacionado con el mundo de los mercados financieros es llegar a entender de qué manera se ven afectados por factores externos los valores de la bolsa y si se puede llegar a predecir su comportamiento, pudiendo hacer más fácil la interacción con la materia a usuarios con nivel bajo/medio en conocimientos de transacciones y mercados de valores para que puedan tener una guía en la toma de decisiones.

Uno de los criterios utilizados a la hora de realizar operaciones en bolsa es la información de los mercados y noticias sobre las empresas relacionadas, que nos ayudan a prever subidas o bajadas. Por ello, se ha utilizado un repositorio de noticias y los valores al cierre de las empresas cotizadas en la Bolsa de Madrid de todo 2014 como base del estudio.

### 1.3 Objetivos del proyecto

Se puede dividir el proyecto en dos partes: la obtención de información y el análisis de los resultados obtenidos.

#### 1.3.1 Objetivo principal

El objetivo principal de este proyecto es crear una herramienta para ayudar al usuario no experto en el dominio financiero a tomar decisiones sobre la compra/venta de acciones en Bolsa.

Para ello se estudia la influencia de las noticias en el mercado de valores analizando el sentimiento de las noticias clasificándolas como positiva, negativa o neutra. Este análisis es realizado con una serie de herramientas obtenidas de manera externa o han sido creadas de forma autónoma.

Como se ha comentado anteriormente se puede dividir en dos fases definidas de la siguiente manera:

- Obtención de datos reales:
  - Valores de cierre diarios en 2014 de las empresas cotizadas en el Ibex35
  - Noticias de ámbito económico en 2014
- Valoración de noticias con analizadores de sentimientos
- Creación de la herramienta, análisis, visualización y conclusiones

### 1.3.2 Objetivos secundarios

Como objetivos secundarios se podrían considerar el estudio de proyectos similares y análisis de posibles mejoras y trabajos futuros.

## 1.4 Estructura del documento

En primer lugar se ofrece un resumen del proyecto tanto en español como en inglés y se dedican unas palabras de agradecimiento a toda la gente que ha hecho posible que el proyecto salga adelante.

A partir del índice del documento, la estructura es la siguiente:

1. Introducción: en ella se realiza un breve prólogo del proyecto, la motivación que me ha llevado a realizarlo, los objetivos y la estructura del documento.
2. Estado del arte: es el punto donde se han invertido más horas de trabajo y más extenso del documento.

Se analizan proyectos similares, el proceso de obtención de los datos, se realiza un estudio en profundidad tanto del *análisis de sentimiento* y herramientas de soporte para analizar las noticias. También se detalla la metodología ágil SCRUM utilizada en proyecto.

3. Análisis, diseño e implementación del sistema: Se analizan los requisitos de usuario y las herramientas utilizadas en el estudio. Se presenta la arquitectura de la aplicación y se detallan cada una de las fases de éste.
4. Pruebas: Se muestran las pruebas realizadas por cada una de las herramientas que componen la herramienta (pruebas funcionales y pruebas no funcionales).
5. Planificación: Se describe el ciclo de vida implementado en el proyecto. También se define la planificación estimada y la planificación real, en ambas se ha diseñado un diagrama de horas y un diagrama de Gantt con las fechas.
6. Presupuesto: Al igual que en la planificación se define un presupuesto inicial, uno final y las diferencias que han existido entre uno y otro.
7. Conclusiones: Se detallan las conclusiones del trabajo y un breve resumen de éste.
8. Trabajos Futuros: Se describen las mejoras que se pueden realizar en el proyecto que por falta de tiempo o de herramientas no ha sido posible implantarlas en esta etapa.
9. Diccionario de Acrónimos: Diccionario con los acrónimos utilizados en el documento.

## CAPÍTULO 1

10. Bibliografía: Se realiza un índice con todos los documentos y los textos que se han utilizado en el documento.
11. ANEXO I: Cálculo Salario Profesionales: se muestran los cálculos realizados para definir el salario de cada profesional.

# Capítulo 2

## Estado del arte

### 2.1 Introducción

En este punto se explica de forma teórica la base sobre la que se ha realizado este Trabajo. En primer lugar se hará un pequeño análisis de proyectos similares, a continuación se describirá brevemente cómo se han extraído los datos iniciales, se analizarán las distintas herramientas de análisis de sentimientos que se han seleccionado inicialmente y finalmente se describe la metodología ágil SCRUM utilizada en el proyecto.

En este apartado se redacta el análisis inicial realizado acerca del estado en que se encuentran en la actualidad las tecnologías utilizadas en este proyecto. Este estudio ha sido desarrollado previamente al diseño de la aplicación y ha servido de guía introductoria y referencia para las siguientes fases del proyecto.

La metodología empleada ha consistido en una primera etapa de investigación puramente teórica en la que se han identificado términos, agrupado conceptos y relacionado temas. Mediante este estudio, se han reforzado las nociones previamente adquiridas acerca de algunos conceptos y se han aprendido otros completamente desconocidos en un principio.

## 2.2 Proyectos similares

Tanto el *Análisis de Sentimientos* como el comportamiento del Mercado de Valores son mundos con infinitud de posibilidades que supone un reto constante para seguir avanzando y descubrir sus misterios. La predicción del comportamiento de los mercados es uno de los retos más buscados por el hombre. A continuación se describen dos proyectos similares y las diferencias sobre el actual proyecto.

### 2.2.1 Estudio de la Universidad de Warwick

Hoy en día existe un gran potencial para hacer revolucionarias predicciones económicas. Así lo afirma, al menos, la todavía incipiente ciencia de los conjuntos masivos de datos, que en su corta vida ya ha cosechado importantes aciertos.

El mercado bursátil es resbaladizo y variable, por lo que su predicción es algo complicada. Aún así, la conclusión de un trabajo realizado por un grupo de investigadores de la Universidad de Warwick, en el Reino Unido, es clara: **es viable** predecir el momento en el que los mercados financieros tenderán a tener una caída.

El grupo de investigadores de la Universidad de Warwick dice haber demostrado que es posible utilizar de forma inteligente los datos que proporcionan la Wikipedia y Google para hacer un análisis conjunto que lleve a los tan ansiados resultados.

Descubrir tendencias, el Santo Grial de los análisis de Big Data, en este caso se traduce en el aprovechamiento de ambos conjuntos de datos aplicando un *modus operandi* que se basa en la aparición de frases en Wikipedia asociadas a anteriores crisis bursátiles, y combinando estos resultados con los términos de búsqueda de Google.

"Utilizamos una técnica muy conocida de la lingüística computacional para identificar temas semánticos dentro de este corpus que es Wikipedia", explican los investigadores. A partir de este punto, contaron con la ayuda de Mechanical Turk para aplicar inteligencia humana y dar nombres a cada una de estas listas obtenidas.

Por último, examinaron la incidencia de estos temas en las búsquedas de Google durante el periodo 2004-2012, - basadas en Google Trend, no en consultas reales, intentando detectar relaciones entre estas búsquedas y los movimientos de los mercados.

La investigación reveló que existe una correlación entre las cosas que las personas buscan en Google antes de tomar una decisión de compra-venta. Ello significa que este tipo de comportamiento podría encontrarse también en otro tipo de escenarios.

Es decir, la investigación sugiere que "las extensiones de estos análisis podrían ofrecer información sobre el flujo de información a gran escala antes de una serie de acontecimientos del mundo real.

"En los eventos complejos, tales como los movimientos de los mercados financieros, la información valiosa puede estar contenida en los datos del motor de búsqueda



mediante palabras clave, pero con conexiones semánticas menos obvias", concluye el estudio.

En otras palabras, todavía es una utopía que los inversores empiecen a usar Google para dar forma a sus estrategias de mercado de valores. Mientras Big Data no demuestre lo contrario de un modo fehaciente, habrá que seguir dándole la razón al señor Warren.

- **Diferencias con el proyecto actual**

El estudio de la Universidad de Warwick se centra en la predicción de crisis bursátiles en base a un estudio de crisis históricas, relacionando los hechos acontecidos y la reacción de los inversores en ese momento. En el proyecto actual se realiza un análisis de sentimiento de una noticia para valorar la influencia en el valor de mercado de la empresa a la que hace referencia la noticia.

## **2.2.2 Estudio de las Universidades de Londres y Boston**

Otro proyecto similar es el experimento llevado a cabo hace unos años por la University College London y Boston University intentando buscar un patrón de inversión de acuerdo con las búsquedas que se hacen en el buscador más popular del mundo.

La estrategia llevada a cabo por la University College London y Boston University batió la rentabilidad del índice norteamericano Dow Jones entre agosto de 2004 y junio de 2011. Para ello, utilizaron la herramienta Google Trends, parte gratuita del famoso buscador que permite medir el volumen de búsquedas de una determinada palabra o conjunto de palabras.

Su estrategia consistió en analizar un grupo de 98 términos tales como "deuda", "acciones", "derivados", "dinero", "ingresos", "paro", "crédito" con dicha herramienta y comparar el número de búsquedas semanales que tenían. Ellos estimaban que las personas buscamos mucho más algo cuando nos preocupa que cuando va bien y que igual ocurriría con los términos económicos anteriores.

De acuerdo con esto, siguieron la siguiente estrategia de inversión:

- 1) Si había disminuido en la última semana respecto a anteriores abrían una posición larga (compraban) en las acciones del índice. En este caso, ganaban dinero si subían.
- 2) Si habían aumentado las búsquedas, las posiciones se vendían a principios de la semana y se volvían a comprar al final de la semana, obteniendo beneficio si el precio había bajado.

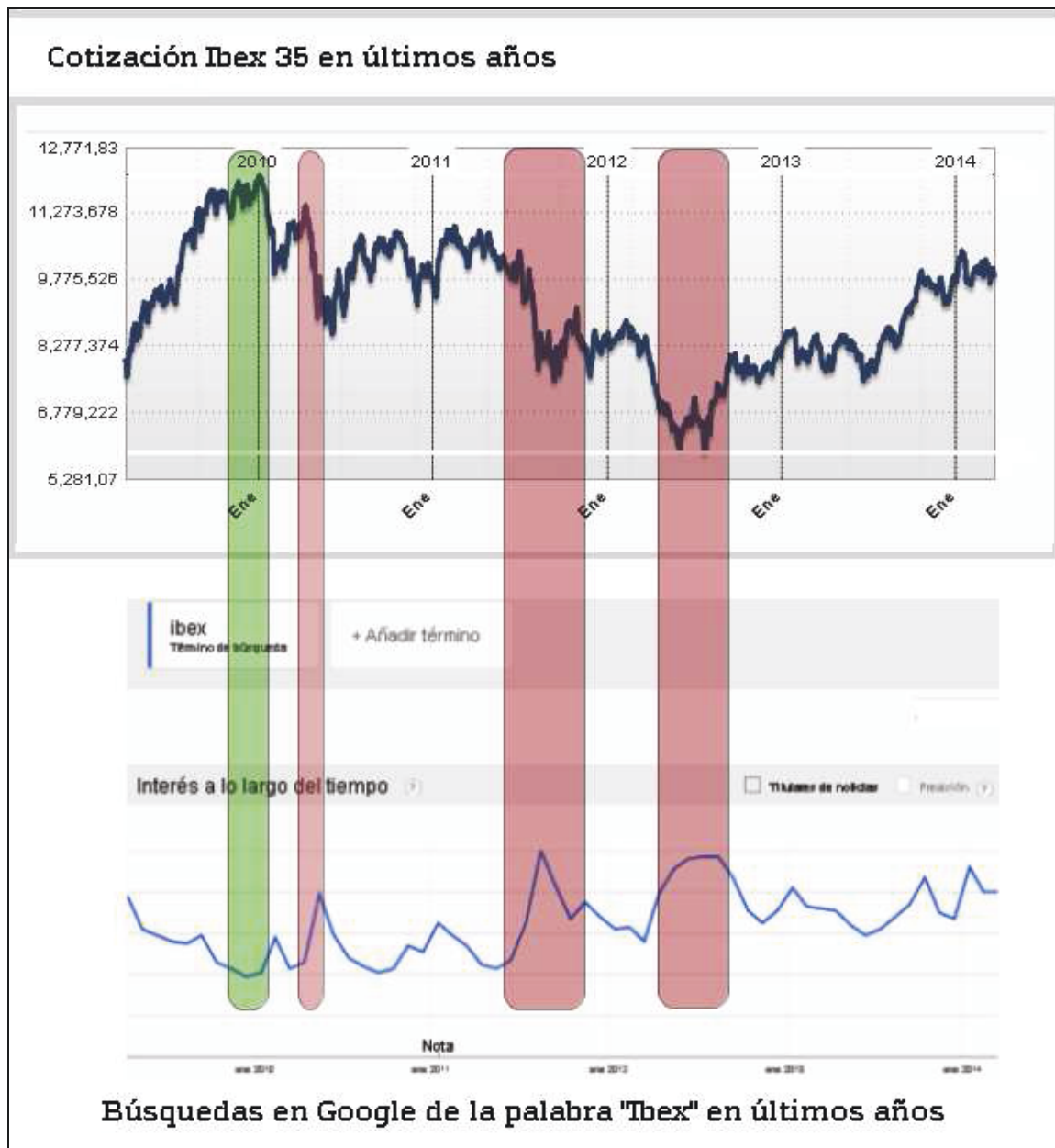
En resumen, su idea era que si había más búsquedas el índice bursátil bajaría porque nos informamos más si las cosas van mal y nos preocupa más perder dinero que ganarlo.

Los resultados del estudio entre 2004 y 2011 ofreció que:

- Rentabilidad del Dow Jones: 16%

- Rentabilidad obtenida con la estrategia utilizada en este caso: 326%

A raíz de este estudio, el equipo de Redes Sociales/Comunicación Online de Bankinter hicieron una adaptación menos elaborada y detallada como los de las universidades anteriores. Compararon el gráfico de la cotización del Ibex 35 con el volumen de búsquedas en Google de la palabra “Ibex” a lo largo de los últimos años.



**Ilustración 1: Gráfico que compara la cotización del Ibex 35 con las búsquedas en Google (Fuente: Blog Bankinter)**

En el mismo se comprueba como quedan especialmente reflejados cambios de tendencia y máximos y mínimos del índice español. Marcando en verde el máximo alcanzado por el Ibex 35, que coincide con el mínimo volumen de búsquedas en el periodo mostrado. En sentido contrario, se observa en las zonas rojas cómo las grandes

caídas y cambios de tendencia del mercado han coincidido con máximos de búsquedas de esta palabra en Google.

- **Diferencias con el proyecto actual**

El estudio de la University College London y Boston University se basa en las búsquedas de los usuarios en Google para, según un patrón, precedir bajadas y subidas en bolsa. Este patrón se centra en palabras clave utilizadas para buscar información a través de un navegador. En este proyecto se utilizan noticias económicas y financieras, no las búsquedas que realiza cualquier usuario, se valorana través de su polaridad para analizar cual es su repercusión. Es decir, se trata con la opinión de expertos y se mide su repercusión.

## 2.3 Fuentes de datos

Para realizar el estudio se necesitan dos tipos de registros de datos: valor al cierre diario para cada empresa cotizada en el Ibex35 y las noticias de índole económico. Se ha decidido utilizar datos reales y una muestra lo suficientemente grande para sacar conclusiones más precisas. Esta información se almacenará en una base de datos en Microsoft Access 2010.

### 2.3.1 Datos Bolsa de Madrid 2014

A través de la web oficial ([www.bolsamadrid.es](http://www.bolsamadrid.es)) se ha obtenido el histórico de valores de cierre de las empresas cotizadas en la Bolsa de Madrid para el año 2014. Estos datos se han almacenado de manera temporal en una hoja Excel con los siguientes campos:

- Fecha: día del cierre de la Bolsa. Por ejemplo, el 02/04/2014 la empresa 'ACS' terminó la jornada con un valor de mercado de 28,385 la acción.
- Empresa: nombre de la empresa cotizada en la Bolsa de Madrid. Las empresas y su código de la Bolsa de Madrid son las siguientes:
  - ABE SM – ABERTIS
  - ABG/P SM – ABENGOA B
  - ACS SM – ACS
  - AMS SM – AMADEUS
  - ANA SM – ACCIONA
  - BBVA SM – BBVA
  - BKIA SM – BANKIA
  - BKT SM – BANKINTER
  - BME SM – BME
  - CABK SM – CAIXABANK

## CAPÍTULO 2

- DIA SM – DIA
  - ELE SM – ENDESA
  - ENG SM – ENAGAS
  - FCC SM – FCC
  - FER SM – FERROVIAL
  - GAM SM – GAMESA
  - GAS SM - GAS NATURAL
  - GRF SM – GRIFOLS
  - IAG SM – IAG
  - IBE SM – IBERDROLA
  - IDR SM – INDRA
  - ITX SM – INDITEX
  - JAZ SM – JAZZTEL
  - MAP SM – MAPFRE
  - MTS SM - MITTAL STEEL COMPANY
  - OHL SM – OHL
  - POP SM – BANCO POPULAR
  - REE SM – RED ELECTRICA
  - REP SM – REPSOL
  - SAB SM – BANCO SABADELL
  - SAN SM – SANTANDER
  - SCYR SM – SACYR
  - TEF SM – TELEFÓNICA
  - TL5 SM – TELE 5
  - TRE SM – TÉCNICAS REUNIDAS
- Descripción: nombre extenso de la empresa cotizada en la Bolsa de Madrid
  - Cierre: valor del cierre diario de la empresa en una fecha

El rango de días almacenado son todos los días de 2014 a excepción de sábados, domingos y los festivos indicados que la Bolsa de Madrid no abre: 1 de Enero, 18 y 21 de Abril, 1 de Mayo, 25 y 26 de Diciembre.

Se ha almacenado 1 registro por cada empresa, día hábil y valor de cierre con un total de 8.925 registros.

30/05/2014	SCYR SM	SACYR	4,972
------------	---------	-------	-------

**Ilustración 2: Ejemplo de un registro del cierre de Bolsa**

### 2.3.2 Noticias de 2014

Las noticias seleccionadas para realizar el estudio se han recuperado de la hemeroteca del diario “Expansión” (<http://www.expansion.com/hemeroteca/>) almacenando un total de 28.630 noticias.

“Expansión es un periódico económico español fundado el 27 de mayo de 1986. El diario es líder en España de la prensa económica con una tirada de 160.000 ejemplares a finales de 2008. Perteneció al grupo Unidad Editorial tras la compra en 2007 por parte de RCS MediaGroup del Grupo Recoletos.” ([Definición Wikipedia](#))

Las noticias se han almacenado de manera temporal en un Excel con los siguientes campos:

- Fecha: día en el que se publicó la noticia
- Titular: pequeña descripción de la noticia
- Noticia: descripción más extensa de la noticia

El rango de días almacenados son todos los días de 2014 a excepción de fines de semana y festivos.

07/01/2014	37. Sacyr, el valor más recomendado por el consenso de mercado pese al conflicto del Canal de Panamá	Sacyr es el valor más recomendado por los analistas de consenso de mercado en enero de 2014, con un precio objetivo de 4,46 euros y un beneficio por acción de 0,32, pese a los bruscos movimientos que ha experimentado en los últimos días en bolsa después de anunciar que podría paralizar las obras del Canal de Panamá.
------------	--	---

**Ilustración 3: Ejemplo de un registro de la tabla Noticia**

## 2.4 Análisis de sentimiento

El *análisis de sentimiento* (conocido también como minería de opinión o *sentimental analysis* en inglés) se define como el uso de procesamiento de lenguaje natural, análisis de texto y lingüística computacional para identificar y extraer información subjetiva de un recurso.

En términos generales, el análisis de sentimiento intenta determinar la actitud de un interlocutor o un escritor con respecto a algún tema o la polaridad contextual general de un documento. La actitud puede ser su juicio o evaluación, estado afectivo (o sea, el

## CAPÍTULO 2

estado emocional del autor al momento de escribir), o la intención comunicativa emocional (o sea, el efecto emocional que el autor intenta causar en el lector).

Una tarea básica en análisis de sentimientos es clasificar la polaridad de un texto dado a nivel de documento, oración, o rasgo/característica — si la opinión expresada en un documento, una oración o un rasgo/característica de una entidad es positiva, negativa, o neutral. La clasificación de sentimiento más avanzada, "más allá de la polaridad" busca, por ejemplo, estados emocionales tales como "enfadado", "triste", y "feliz".

A los efectos de este proyecto, "la detección de sentimiento" significa encontrar la polaridad (positiva, negativa o neutra) de un texto dado.

Actualmente las redes sociales son el ámbito donde más se ha investigado y desarrollado sobre este tema, y muchas aplicaciones se han centrado en esta rama, aún así es extrapolable a otros campos. Vamos a analizar varios analizadores de sentimientos sin tener en cuenta si es exclusivo para analizar textos en redes sociales.

Para la evaluación, se utilizaron herramientas comerciales para la detección automática sentimiento. Existen literalmente cientos de estas herramientas. Con el fin de obtener resultados comparables, las herramientas tenían que cumplir con los siguientes requisitos: herramienta independiente de detección de sentimiento (es decir, no forman parte de un sistema mayor, como por ejemplo los sistemas de monitorización de medios sociales); capacidad para analizar textos arbitrarios (es decir, no especializados en tipos de texto individuales como los tweets); Acceso a la API; el acceso sin coste.

### 2.4.1 Semantria

Semantria es una herramienta de la empresa Lexalytics, de la que se escindió en 2011. Semantria ofrece análisis de texto a través de API's y un plugin de Excel. Incorpora una gran base de conocimientos y utiliza deep learning (análisis profundo).

Aprendizaje profundo (en inglés, deep learning) es un conjunto de algoritmos en aprendizaje automático (en inglés, machine learning) que intenta modelar abstracciones de alto nivel en datos usando arquitecturas compuestas de transformaciones no-lineales múltiples.

Aprendizaje profundo es parte de un conjunto más amplio de métodos de aprendizaje automático basados en aprender representaciones de datos. Una observación (por ejemplo, una imagen) puede ser representada en muchas formas (por ejemplo, un vector de píxeles), pero algunas representaciones hacen más fácil aprender tareas de interés (por ejemplo, ¿es esta imagen una cara humana?) en base a ejemplos, y la investigación en esta área intenta definir cuáles de las representaciones son mejores y cómo crear modelos para aprender estas representaciones.

Varias arquitecturas de aprendizaje profundo, como redes neuronales profundas, redes neuronales profundas convolucionales, y redes de creencia profundas, han sido aplicadas a campos como visión por computador, reconocimiento automático del habla, y

reconocimiento de señales de audio y música, y han mostrado producir resultados de vanguardia en varias tareas.

- **¿Cómo se accede a Semantria?**

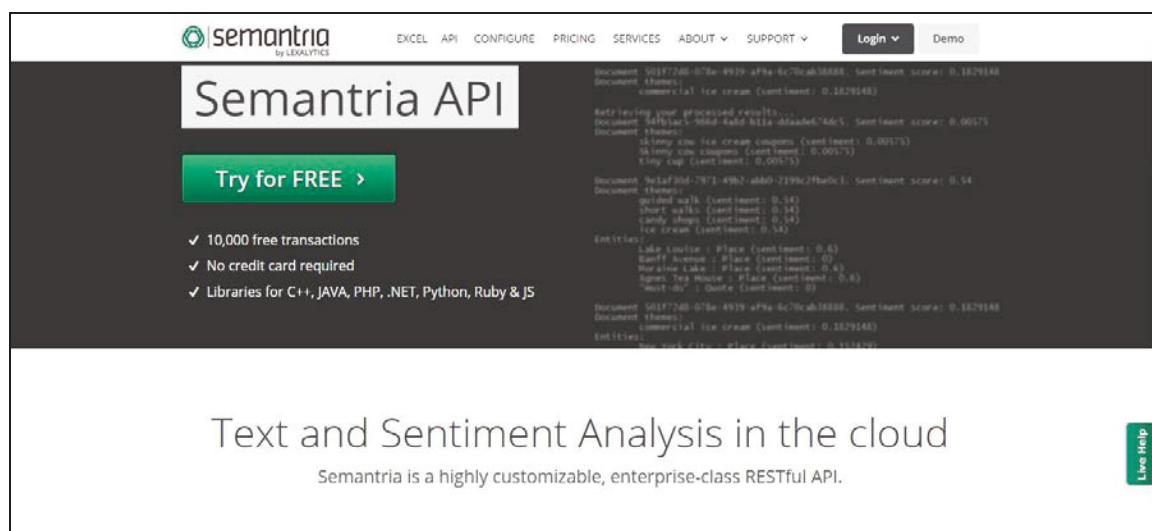
Para poder utilizar el API o el plugin de Excel de Semantria es necesario registrarse en la web de Semantria. Para ello sólo hay que inscribirse dentro de esta web <https://semantria.com/signup>.

- **¿Gratuito o de pago?**

Tiene una versión de prueba de 30 días con una limitación de 20.000 análisis.



**Ilustración 4: Página web principal de Semantria for Excel**  
<https://semantria.com/excel>

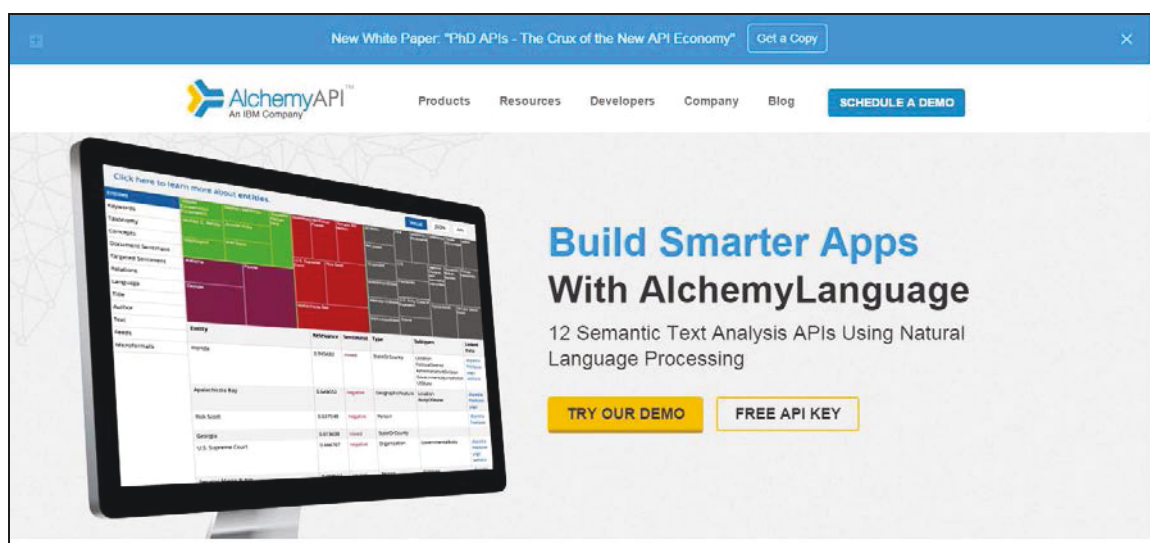


**Ilustración 5: Página web principal de Semantria API**  
<https://semantria.com/API>



## 2.4.2 Alchemy API

AlchemyAPI es una empresa que utiliza aprendizaje automático (en concreto, el aprendizaje profundo) para hacer el procesamiento de lenguaje natural (en concreto, el análisis de textos semántica, incluyendo análisis de los sentimientos) y la visión por ordenador (en concreto, la detección y reconocimiento de rostros) para sus clientes, tanto sobre la nube y on-premises. A partir de febrero de 2014, dice tener clientes en 36 países y procesado más de 3 billones de documentos al mes. ProgrammableWeb añadió AlchemyAPI a su API Billionaires Club (API Request con más de 1 billón de peticiones diarias) en septiembre de 2011.



**Ilustración 6: Página web principal de AlchemyAPI**  
<http://www.alchemyapi.com/>

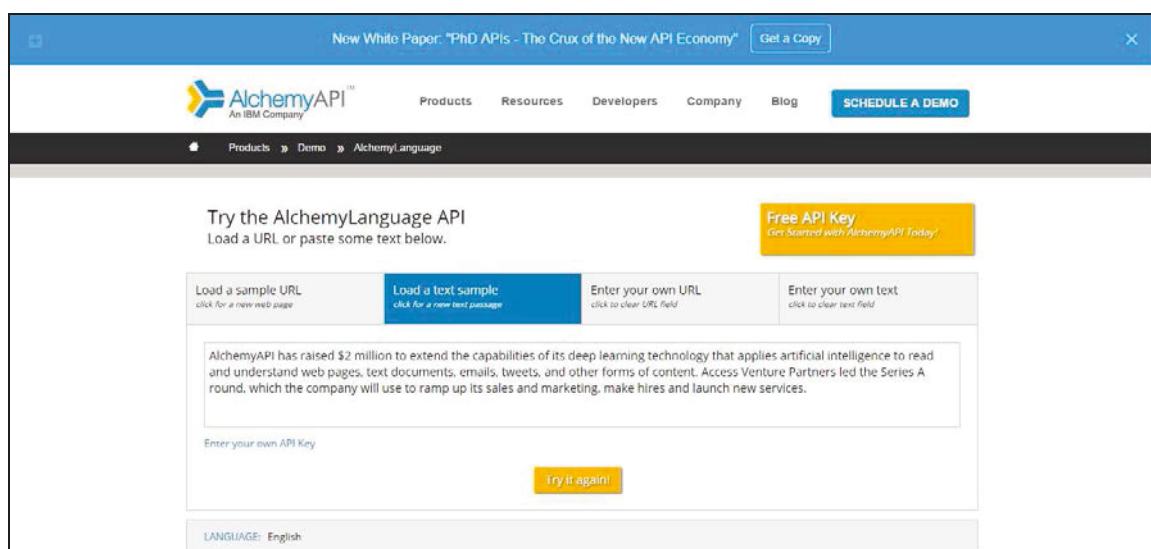
- **¿Cómo se accede a AlchemyAPI?**

Para poder utilizar el API de AlchemyAPI es necesario registrarse en la web de AlchemyAPI. Para ello sólo hay que inscribirse dentro de esta web <http://www.alchemyapi.com/>.

- **¿Gratuito o de pago?**

Tiene una versión gratuita con una limitación de 20.000 análisis y una versión demo en formato web para probar puntualmente tanto urls como texto plano.



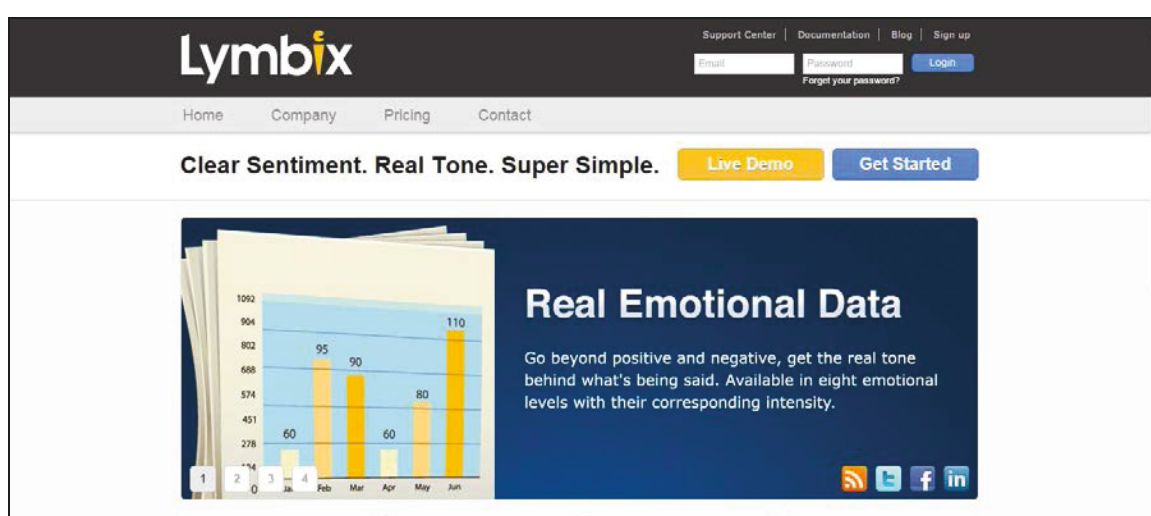


**Ilustración 7: Página web AlchemyAPI. Demo web**

## 2.4.3 Lymbix

En 2009, Matt Eldridge (ahora miembro del consejo) y Josh fundaron Lymbix con la idea de *ToneCheck*™ que ayuda a la gente a evitar flamewars de correo electrónico. El desarrollo de *ToneCheck*™ fue el catalizador para la comprensión innovadora de análisis de los sentimientos, el análisis del tono, la extracción de entidad que no sólo compite en el área de análisis de texto.

Vieron una necesidad en el mercado. Los expertos de una variedad de industrias clamaban por la claridad, la escalabilidad y la capacidad de adaptación de un lenguaje en constante evolución. La metodología de combinar la PLN con el aprendizaje adaptativo se reunieron a partir de técnicas de propulsión humana ha dado lugar a un híbrido construido para el análisis de la web social.



**Ilustración 8: Página web principal de Lymbix. <http://www.lymbix.com/>**

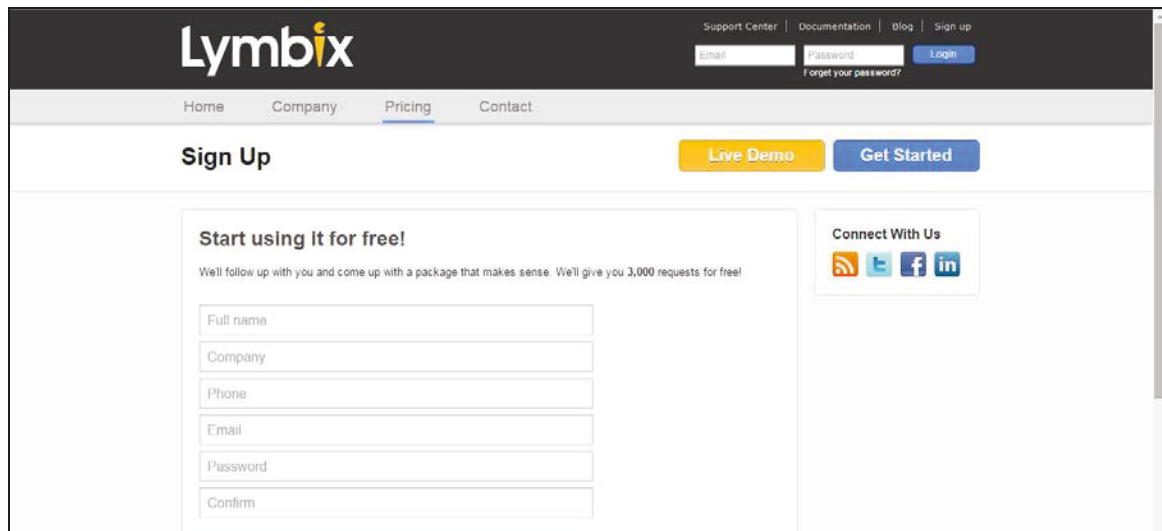
- ¿Cómo se accede a Lymbix?

## CAPÍTULO 2

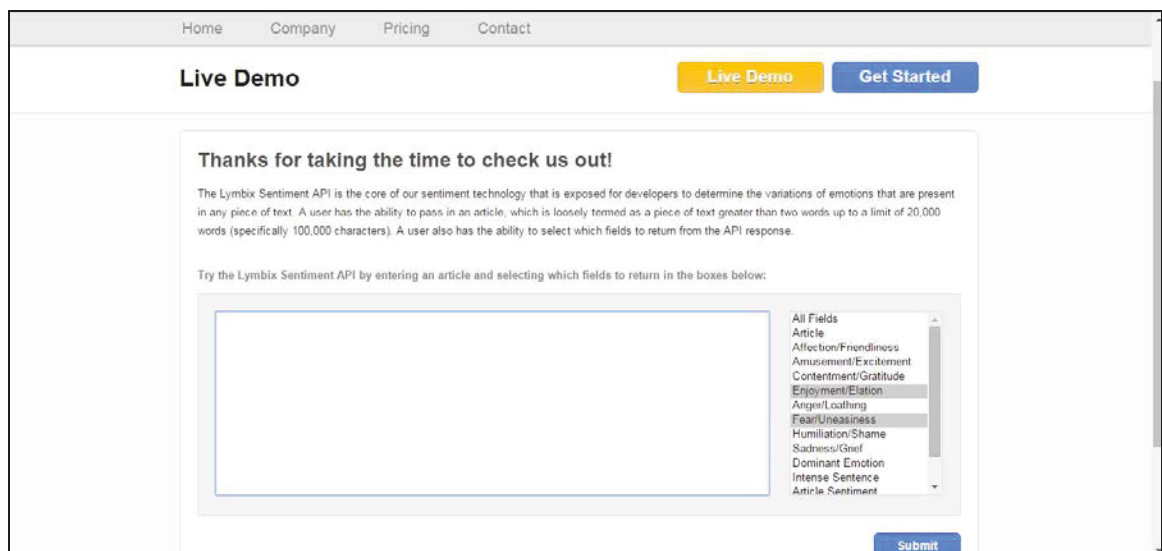
Para poder utilizar el API de Lymbix es necesario registrarse en la web de Lymbix. Para ello sólo hay que inscribirse dentro de esta web <http://www.lymbix.com/users/new>.

- **¿Gratis o de pago?**

Tiene una versión gratuita con una limitación de 3.000 request y una versión demo en formato web para probar texto plano con un cuadro de texto y un listado de tipos de texto.

The screenshot shows the Lymbix website's sign-up page. At the top, there's a navigation bar with links for Support Center, Documentation, Blog, and Sign up. Below this is a header with the Lymbix logo and a login section with fields for email and password, and a 'Login' button. The main content area is titled 'Sign Up' and features a 'Live Demo' button and a 'Get Started' button. A central box titled 'Start using it for free!' contains a form with fields for Full name, Company, Phone, Email, Password, and Confirm. To the right of the form is a 'Connect With Us' section with social media icons for RSS, Twitter, Facebook, and LinkedIn.

**Ilustración 9: Página web Lymbix. Sign Up free trial**

The screenshot shows the Lymbix website's live demo page. At the top, there's a navigation bar with links for Home, Company, Pricing, and Contact. Below this is a header with the Lymbix logo and a 'Live Demo' button and a 'Get Started' button. The main content area is titled 'Live Demo' and features a 'Live Demo' button and a 'Get Started' button. A central box titled 'Thanks for taking the time to check us out!' contains a text area for entering an article and a dropdown menu for selecting fields to return from the API response. The dropdown menu lists various emotions and sentiments, including All Fields, Affection/Friendliness, Amusement/Excitement, Contentment/Gratitude, Enjoyment/Elation, Anger/Loathing, Fear/Uneasiness, Humiliation/Shame, Sadness/Grief, Dominant Emotion, Intense Sentence, and Article Sentiment. A 'Submit' button is located at the bottom right of the form.

**Ilustración 10: Página web Lymbix. Live Demo**

### 2.4.4 Repustate

Primero vamos a resumir la metodología que usa la herramienta que se divide en varias fases:

- **Etiquetado gramatical:**

. Antes de calcular cualquier sentimiento, un bloque de texto tiene que ser desglosado en sus construcciones gramaticales. Esto ayuda a guiar algoritmos de aprendizaje automático en la dirección correcta.

La descomposición de texto en sus partes gramaticales se llama part-of-speech tagging. Por ejemplo:

**Realmente me encantó la pizza anoche**

Ejecutando el part-of-speech tagger de Repustate devuelve la siguiente salida:

**Realmente(ADV) me(PP) encantó(VB) la(ART) pizza(NN) anoche(NN)**

Después de cada palabra, el "tag" apropiado le sigue. ADV significa adverbio, PP significa nombre propio, VB significa verbo y así sucesivamente. Un conjunto de etiquetas completo está disponible en su web <https://www.repustate.com/docs/>. Con las etiquetas aplicadas, ahora se tiene una pista sobre qué palabras o frases pueden ser de interés. Por ejemplo, ART se refiere a un artículo. Sabemos artículos nunca influyen en una sentencia, por lo que se pueden ignorar estas palabras.

- **Lematización:**

Un paso final antes de conseguir el sentimiento; el concepto de lematización de una palabra. En todos los idiomas, la misma palabra puede escribirse de manera diferente en función de varios factores, entre ellos tiempo, género y número. Algunos lenguajes, como el árabe, tienen reglas más complejas, mientras que en inglés, hay menos reglas que preocuparse como inglés no tiene un concepto de género para las palabras. Vamos a ilustrar esto con un ejemplo:

**I really enjoyed that movie**

**I really enjoy watching movies**

Para un ser humano, es obvio que estas frases son similares, pero estamos tratando de enseñar a una máquina cómo saber eso. La primera cosa a tener en cuenta es que *enjoy* y *enjoyed* son similares. Cuando se lematiza la palabra *enjoyed*, es decir, se reduce a su forma de raíz, obtienes **enjoy**. Lo mismo para las *movies* y *movie*. Se deben lematizar palabras con el fin de simplificar el análisis de los sentimientos, ya que reduce el número total de palabras que hay que considerar cuando se busca la polaridad.

- **Prior Polarity (Polaridad anterior) u Orientación semántica:**

En su forma más simple, análisis de sentimientos se trata de encontrar las palabras individuales que tienen el sentimiento asociado con ellos. Palabras como "amor", "odio", "grande", "terrible", etc. transmiten de inmediato un sentimiento particular, o la polaridad

## CAPÍTULO 2

("polaridad", que significa el sentimiento positivo o negativo). Estos términos se dice que tienen una polaridad anterior.

¿Pero es el análisis de sentimiento realmente es sólo una cuestión de encontrar términos de polaridad anteriores? Si es así, entonces hemos terminado aquí. Pero, por supuesto no lo es y hay mucho más que eso. Considere este ejemplo:

### **I did not like that restaurant last night**

La palabra *like*, como verbo, transmite sentimiento positivo, pero todo el mundo estaría de acuerdo en que esta sentencia es negativa en lo que respecta al restaurante. O ¿qué pasa con esta frase?:

### **That dinner could not have been any better**

Como estos dos ejemplos muestran, términos de polaridad anteriores por sí solos no son suficientes. Se necesita más análisis para determinar con precisión el sentimiento.

- **Negaciones y Amplificadores**

Una negación invierte la polaridad del siguiente(o, a veces, anterior) término. Al igual que en el ejemplo anterior, *did not like* en realidad es negativa debido a la presencia de *not* en precedencia a *likes*. La frase *could not have been* es lo que llamamos un "amplificador". A pesar de que contiene una negación, lo que realmente hace es amplificar el término o frase que le sigue. Hay que tener en cuenta que un término positivo o negativo puede seguir esta este particular amplificado.

Negaciones y amplificadores hacen el análisis de los sentimientos más complejo, ya que significa simplemente buscar términos de polaridad anteriores y luego sumando las ocurrencias o algo similar ingenioso simplemente no va a funcionar.

- **Sentimiento de documento**

Al calcular el sentimiento de todo un bloque de texto, esto se llama el *sentimiento de documento*. A menudo, esta puntuación es suficiente. Al examinar pequeños fragmentos de texto, como el contenido de Twitter, por lo general se puede conseguir análisis de sentimiento del documento. Editoriales de los sitios de noticias o artículos del blog también, en general se valen de ser analizadas en su totalidad. Así que, ¿cuando falle el sentimiento de documento? Considerese este bloque de texto:

**We stayed at the Ritz for 4 days last month. The food and service was exquisite, but the price was way too expensive compared to surrounding hotels**

Este fragmento de opinión de un hotel empieza como algo positivo, pero luego se desvía en la dirección negativa. Analizando este documento en su totalidad puede ocultar este matiz. Si usted es el director del hotel, es posible que desee saber qué aspectos del hotel la gente disfruta y cuáles no.

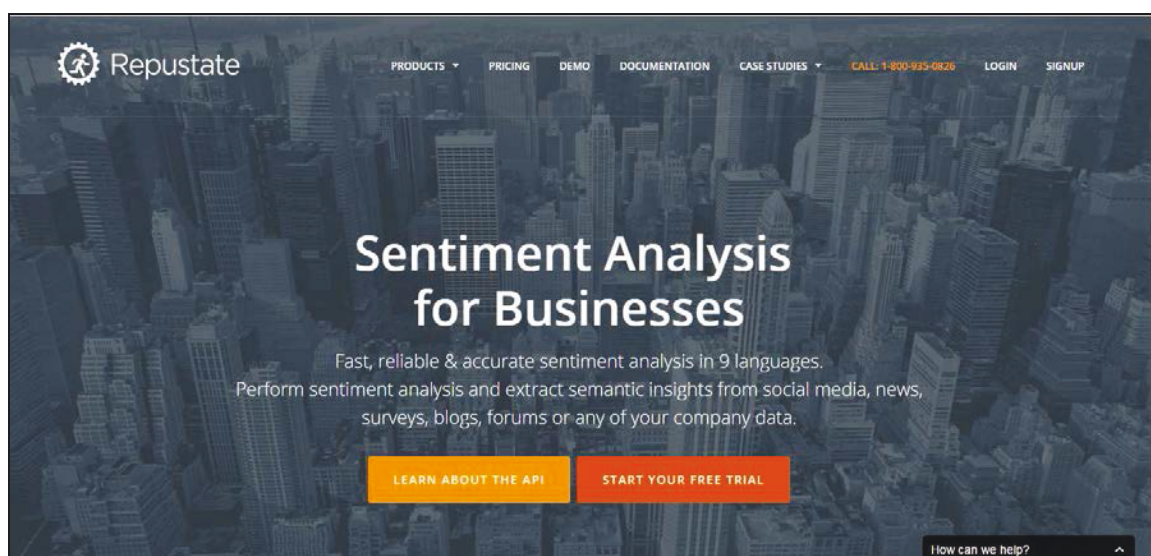
- **Sentimiento de ámbito**

Para combatir el problema de las múltiples polaridades que aparecen dentro de un bloque de texto, el API de Repustate permite el sentimiento de ámbito. Esto significa que puede obtener múltiples puntuaciones de sentimiento desde el mismo bloque de texto que están en el ámbito de un tema en particular (por ejemplo, una persona o un lugar) o de ámbito a la idea semántica (por ejemplo, la fijación de precios en un hotel). Al utilizar el sentimiento de ámbito, el matiz que existe en una revisión se lleva hasta la superficie y todos los sentimientos relevantes son capturados.

- **Sentimiento personalizado**

Diferentes industrias emplean diferentes medios de comunicación. Palabras o frases que tienen un solo connotación en un contexto podrían tener una completamente diferente en otro. Es casi imposible darse cuenta de todas estas variaciones en un modelo universal. Repustate permite crear reglas sentimiento que son específicas de una industria específica y de casos de uso.

Análisis multilingüe de sentimiento es difícil, especialmente cuando se trata de texto en un contexto real. Repustate ha desarrollado múltiples técnicas que cuando se ponen a utilizar juntos, ofrecen un rápido, preciso y fiable motor de análisis de texto.



**Ilustración 11:** Página web principal de Repustate. <http://www.repustate.com/>

- **¿Cómo se accede a Repustate?**

Para poder utilizar el API de Repustate es necesario registrarse en la web de Repustate. Para ello sólo hay que inscribirse dentro de esta web <https://www.repustate.com/signup/>.

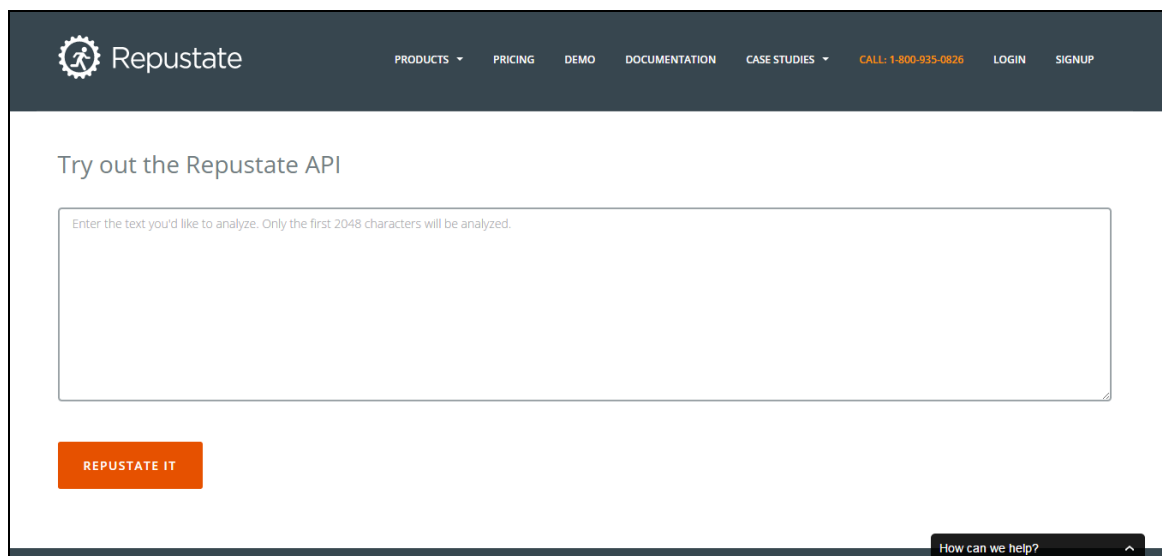
- **¿Gratuito o de pago?**

Tiene una versión gratuita con un periodo de prueba de 14 días y una versión de pago con tres opciones distintas:

- Developer:

- \$79 mensuales
  - 10.000 documentos mensuales
  - Análisis semántico y sentimental
  - Soporta todos los idiomas
  - Soporte por email
  - Resultados de sentimientos personalizados
- Business:
    - \$499 mensuales
    - 100.000 documentos mensuales
    - Análisis semántico y sentimental
    - Soporta todos los idiomas
    - Soporte por email y teléfono
    - Resultados de sentimientos personalizados
    - Sesiones de tutoriales one-on-one
  - Enterprise:
    - \$499 mensuales
    - 1.000.000 documentos mensuales
    - Análisis semántico y sentimental
    - Soporta todos los idiomas
    - Ingeniero de soporte dedicado
    - Resultados de sentimientos personalizados
    - Sesiones de tutoriales one-on-one
    - Acceso prioritario a nuevas características

También dispone de una versión demo on-line donde analizar textos puntuales:



The screenshot shows the Repustate website's demo interface. At the top is a dark navigation bar with the Repustate logo on the left and links for PRODUCTS, PRICING, DEMO, DOCUMENTATION, CASE STUDIES, a phone number (CALL: 1-800-935-0826), LOGIN, and SIGNUP on the right. Below the navigation bar, the main content area has the heading "Try out the Repustate API". Underneath this heading is a large, empty text input box with a placeholder text that reads: "Enter the text you'd like to analyze. Only the first 2048 characters will be analyzed." Below the text box is an orange button labeled "REPUSTATE IT". In the bottom right corner of the main content area, there is a small link that says "How can we help?" with an upward-pointing arrow icon.

**Ilustración 12: Página web Repustate. Demo on-line**



## 2.4.5 MeaningCloud Add-in para Excel

En el proyecto actual nos decantamos por la herramienta Meaning Cloud por varios motivos:

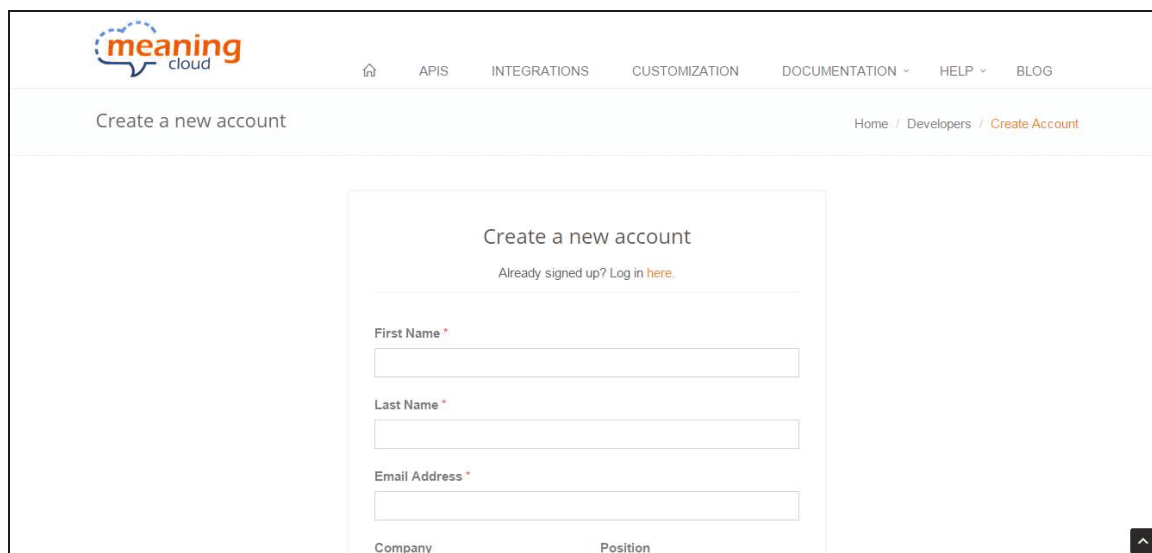
- Disponible add-in para Excel. No necesitamos integrar un API en la herramienta para hacer el estudio puntual de las noticias de 2014
- Mejor integración con la herramienta desarrollada en Access VBA
- Herramienta gratuita
- Extracción rápida y eficiente en un alto volumen de datos



**Ilustración 13: Página web principal de MeaningCloud.**

<https://www.meaningcloud.com/es/>

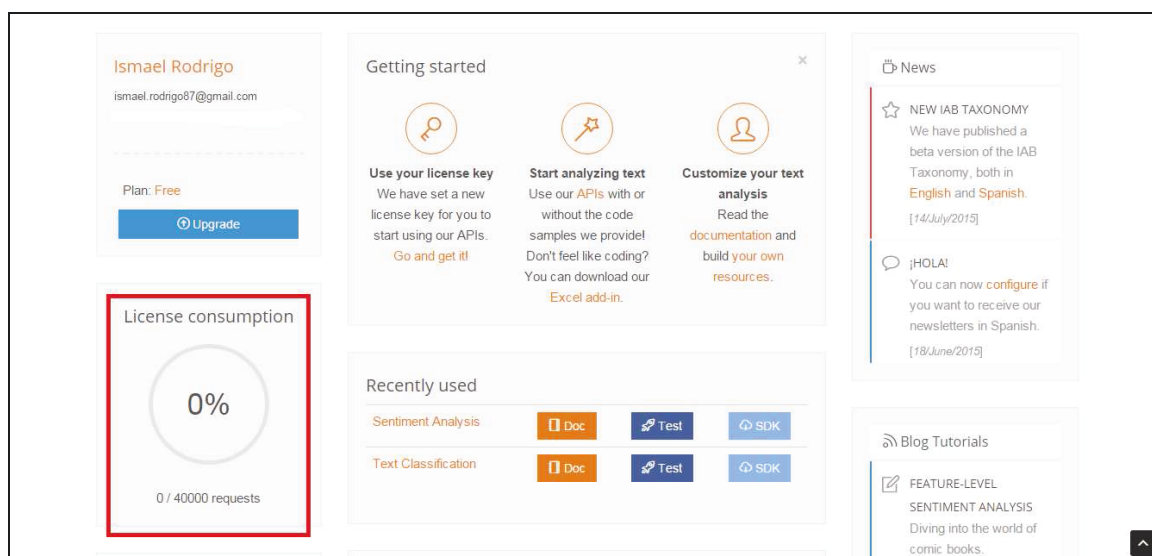
Tras descargar e instalar necesitamos una cuenta. Nos registramos y se nos proporciona una clave para introducir en el Add-in y comenzar a analizar.



The screenshot shows the 'Create a new account' page on the MeaningCloud website. The page has a header with the MeaningCloud logo and navigation links: Home, APIS, INTEGRATIONS, CUSTOMIZATION, DOCUMENTATION, HELP, and BLOG. Below the header, there's a sub-header with 'Create a new account' and a breadcrumb trail: Home / Developers / Create Account. The main content area contains a form with the following fields: First Name, Last Name, Email Address, Company, and Position. There is also a link for 'Already signed up? Log in here.' and a small 'Create Account' button at the bottom right.

**Ilustración 14: Página web de registro en MeaningCloud**

La desventaja es la limitación, 40.000 request mensuales. Podemos hacer el seguimiento de estos en la web de MeaningCloud.



**Ilustración 15: Contador de request de MeaningCloud**

- **¿Cómo se accede a MeaningCloud?**

Para poder utilizar el API o el add-in de MeaningCloud es necesario registrarse en la web de MeaningCloud. Para ello sólo hay que inscribirse dentro de esta web <https://www.repustate.com/signup/>.

- **¿Gratuito o de pago?**

Tiene 5 versiones:

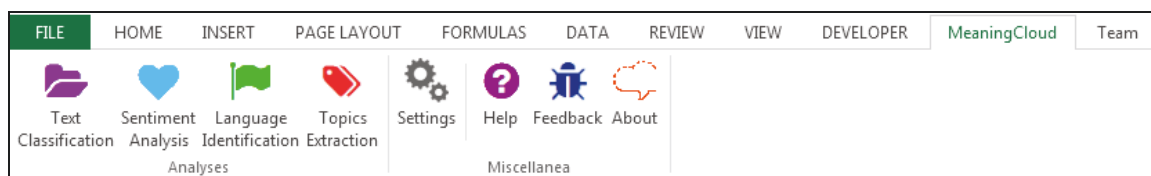
- Free (\$0/mes):



- 40.000 peticiones mensuales
- 2 peticiones por segundo
- Soporte free
- Start-Up (\$99/mes):
  - 120.000 peticiones mensuales
  - 5 peticiones por segundo
  - Soporte standard
- Professional (\$499/mes):
  - 700.000 peticiones mensuales
  - 10 peticiones por segundo
  - Soporte standard
- Start-Up (\$999/mes):
  - 4.200.000 peticiones mensuales
  - 15 peticiones por segundo
  - Soporte standard
- Enterprise (\$ A medida/mes):
  - A medida peticiones mensuales
  - A medida peticiones por segundo
  - Soporte premium

### 2.4.5.1 Interfaz de usuario

En esta sección se explica en detalle las diferentes características incluidas en este complemento. Los análisis que se pueden llevar a cabo, así como las interfaces de usuario que participan tanto en el proceso de análisis y en su configuración.



**Ilustración 16: Ribbon MeaningCloud Add-in Excel**

Está dividido en dos secciones:

- **Analyses**, con los análisis que pueden llevarse a cabo sobre los datos.
  - **Text Classification**, se integra la funcionalidad proporcionada por la Text Classification API. Permite seleccionar qué modelo utilizar para

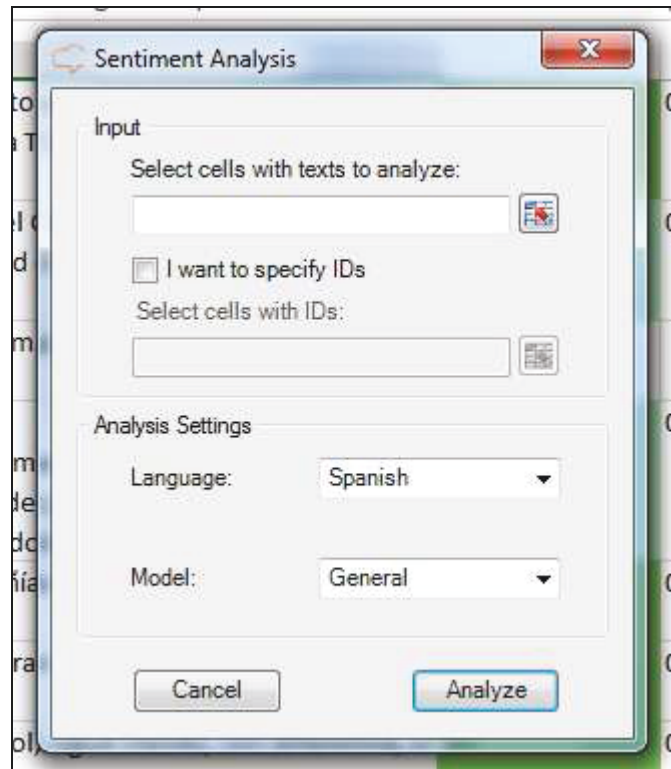
clasificar los textos elegidos, incluidos los modelos definidos por el usuario.

- **Sentiment Analysis**, se integra la funcionalidad proporcionada por la Sentiment Analysis API. Se extrae de un texto dado su polaridad y la subjetividad; la polaridad se proporciona en dos formatos, un valor numérico y una etiqueta.
  - **Language Identification**, se integra la funcionalidad proporcionada por la Language Identification API. Detecta el idioma del texto analizado.
  - **Topics Extraction**, se integra la funcionalidad proporcionada por la Topics Extraction API. Extrae los topics (temas) de los diferentes tipos de texto analizado.
- **Miscellanea**, con la configuración, links relevantes e información.
    - **Settings**, configuración del add-in. Incluye la configuración de clave de licencia, ajustes generales que se aplican a todos los análisis y las opciones específicas para cada uno de los análisis.
    - **Help**, link a documentación online.
    - **Feedback**, link a un formulario de contacto para informar de errores.
    - **About**, información básica sobre el add-in, incluyendo qué versión es.

### 2.4.5.2 Sentiment Analysis

En este análisis se integra parte de la funcionalidad proporcionada por el Sentiment Analysis API. Proporciona dos niveles de análisis para el usuario: un análisis de los sentimientos global, más general de todo el texto y un análisis a nivel de función, donde se extraen las entidades y conceptos y se obtiene su polaridad agregada en el texto.

Esta es la interfaz que aparece al hacer clic en el botón Sentiment Analysis:



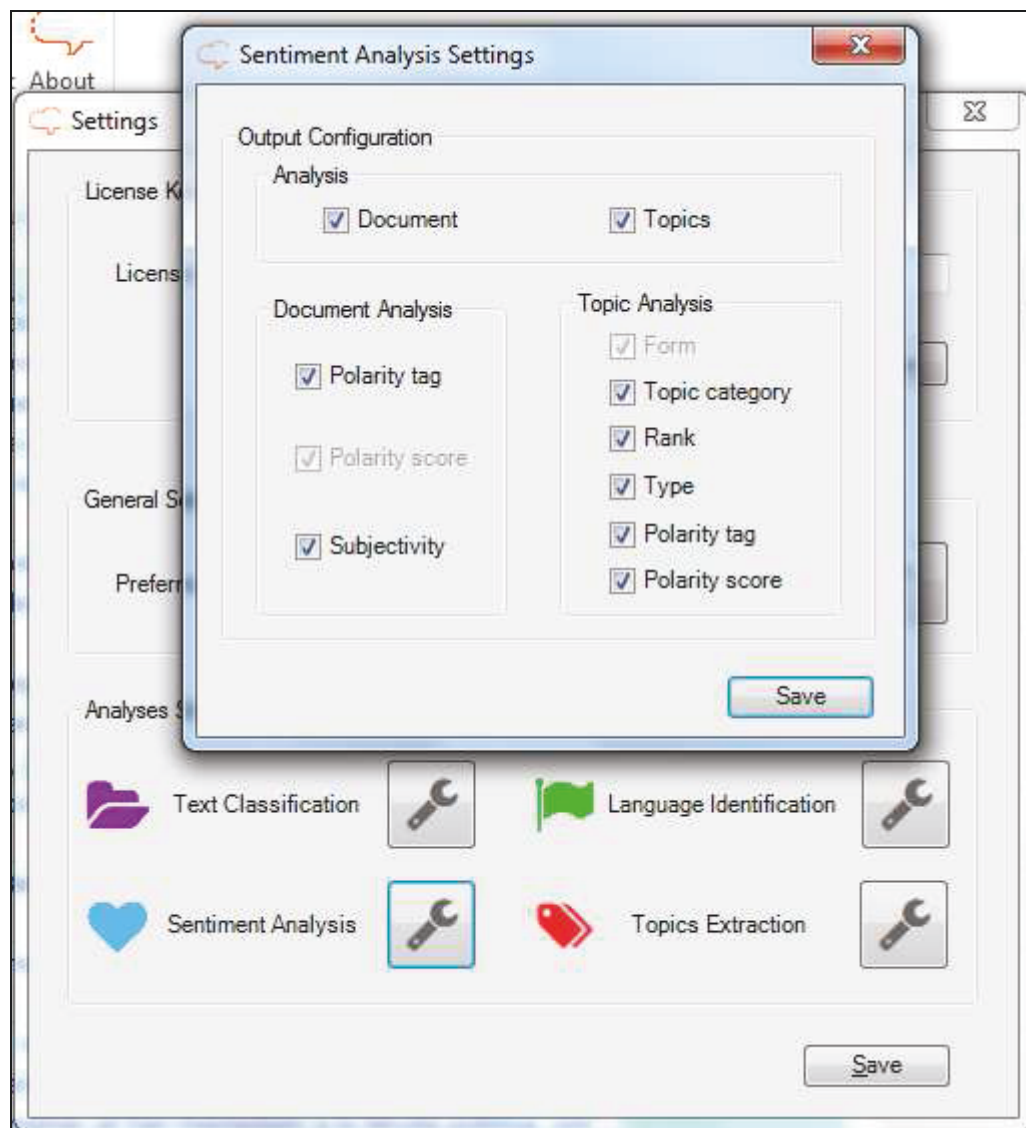
**Ilustración 17: Input Sentiment Analysis**

Se puede ver que hay dos áreas en la interfaz: Input, para seleccionar las celdas que queremos analizar, y Analysis Settings.

En Analysis Settings hay dos valores para seleccionar:

- Language, para seleccionar el idioma de los textos. Por defecto, se preselecciona el idioma establecido como preferido en las Opciones generales. Los valores posibles son: español, inglés y francés.
- Model, con el modelo que se utiliza para medir el sentimiento de los textos. Los modelos de la lista son determinados por el idioma seleccionado en el menú anterior. Actualmente, el único modelo disponible para cada idioma es el general.

Hemos visto en la sección Settings que hay un menú de configuraciones avanzadas con opciones de configuración adicionales para el Sentiment Analysis. Estas son las opciones para el Sentiment Analysis y sus valores por defecto:



**Ilustración 18: Advanced Settings. Sentiment Analysis**

Hay tres aspectos diferentes de Output Configuration que podremos personalizar:

- Analysis: donde podemos seleccionar el tipo de sentiment analysis que queremos llevar a cabo. Hay dos valores posibles, y al menos uno de ellos debe ser seleccionado:
  - *Document*, para extraer el análisis de los sentimientos global de un texto.
  - *Topics*, para extraer los topics (temas) de un texto y encontrar la polaridad asociada a cada uno de ellos.

Ambos análisis son independientes el uno del otro y se seleccionan de forma predeterminada.

- Document Analysis: en esta sección se pueden configurar los campos que queremos dar salida al analizar el sentimiento de un documento. Hay tres posibilidades:

- Polarity tag: muestra la etiqueta de polaridad derivada de la puntuación de polaridad obtenida para el texto. Si no se detecta polaridad, su valor será NONE; cada valor tendrá asociado uno de los siguientes colores:

Polarity Tag	
P+	
P	
NEU	
NONE	
N	
N+	

**Ilustración 19: Polarity Tags**

- Polarity score: muestra la puntuación obtenida por la polaridad del texto; no es configurable, por lo que siempre aparece en los resultados.
- Subjectivity: muestra el valor obtenido para la subjetividad del texto.
- Topic Analysis: en esta sección se pueden configurar los campos que queremos dar salida al analizar el sentimiento de los topics en un documento. Hay seis opciones:
  - Form: con el nombre con el que se identifica el topic extraído. No es configurable, por lo que siempre aparece en los resultados.
  - Topic category: muestra el tipo de topic extraído, es decir, si se trata de una entidad o un concepto.
  - Rank: contiene el orden en que se han detectado los topics. Es específico para cada tipo de topic, es decir, la primera entidad detectada será clasificada con un 1, y el primer concepto se será clasificada con 1 también.
  - Type: muestra el tipo asociado al topic de acuerdo a nuestra ontología.
  - Polarity tag: muestra la etiqueta de polaridad derivada de la puntuación de polaridad obtenida para el topic. Se comporta del mismo modo que la etiqueta de polaridad del Document Analysis.
  - Polarity score: muestra la puntuación de polaridad obtenida para el topic; su comportamiento es el mismo que la puntuación de polaridad visto para Document Analysis, aunque en este caso se puede configurar si aparece en la salida o no.

Hay dos tipos diferentes de análisis de sentimientos que pueden realizarse. Ambos se mostrarán en una nueva hoja de Excel:

- El análisis de nivel de documento se mostrará en una hoja llamada " Global Sentiment Analysis".

## CAPÍTULO 2

- El análisis a nivel de funciones se mostrará en una hoja llamada "Topics Sentiment Analysis".

Esto significa que si se configuran ambos análisis, Sentiment Analysis dará lugar a dos nuevas hojas de Excel.

Ambas hojas incluirán una columna con el texto original, una columna con los ID (si se activa), y luego una columna para cada uno de los campos de salida configurados en la configuración avanzada.

Este es un ejemplo de una posible salida de la polaridad de nivel de documento de textos en español utilizando el modelo general. Está configurado para mostrar todos los campos posibles de salida disponibles con la opción IDs deshabilitada:

	A	B	C	D
	Text	Polarity Tag	Polarity Score	Subjectivity
1	Las tarifas de algunos servicios básicos como la luz, el billete de tren o los precios de las autopistas vuelven a subir con el inicio del nuevo año.	N	-0,6	OBJECTIVE
2	El ministro de Economía, Luis de Guindos, ha asegurado que en 2014 habrá creación neta de empleo que será "incluso superior" a la proyectada en el mes de septiembre cuando se elaboraron los Presupuestos Generales del Estado.	P	0,6	OBJECTIVE
3	Seis de las ocho mayores suspensiones de pagos del año las protagonizaron empresas vinculadas al negocio del ladrillo, como Fergo Aisa, Renta Corporación, Amrey y Contratas y Obras.	N+	-1	OBJECTIVE
4	El Grupo Unidos por el Canal (GUPEC), el consorcio encargado de la ampliación del Canal de Panamá liderado por la española Sacyr, ha anunciado que suspenderá las obras debido a la imposibilidad de llevarlas a cabo con los costos pactados inicialmente en el contrato.	P	0,6	OBJECTIVE
5	El grupo europeo de aeroespacial y defensa EADS pasa a partir de hoy, en el marco de una reestructuración de sus diferentes ramas, a llamarse Airbus, el nombre de la filial que constituye el grueso de su actividad, y que podría haber cerrado 2013 con un nuevo récord de venta de aviones.	NEU	0,1	OBJECTIVE
6	Viene de Suecia, se llama Luna Foreo y es lo último en 'gadgets' de belleza. Tanto para ella como para él, el revolucionario sistema de limpieza facial con filamentos de silicona y tecnología T-Sonic resulta una verdadera solución 'after-party' y un gran propósito para cuidarse el Año Nuevo.	P+	0,64	OBJECTIVE
7	El Grupo Unidos por el Canal (GUPEC), el consorcio encargado de la ampliación del Canal de Panamá liderado por la española Sacyr, ha anunciado que suspenderá las obras debido a la imposibilidad de llevarlas a cabo con los costos pactados inicialmente en el contrato.	P	0,6	OBJECTIVE
8	El fabricante italiano de automóviles Fiat, que controla un 58,5% de Chrysler, comprará en 100% de la compañía norteamericana por 3.650 millones de dólares (2.647,4 millones de euros).	NONE		OBJECTIVE

**Ilustración 20: Ejemplo Global Sentiment Analysis**

El siguiente ejemplo mostrará una posible salida del análisis de los sentimientos en un nivel de funciones, que mostrará los topics que se encuentran en el texto analizado y también si cada uno de esos topics tiene asociado una polaridad. El análisis del ejemplo se ha realizado en español, utilizando el modelo general, y ajustar todos los posibles campos disponibles para ser mostrado en la salida. En este caso, no hemos seleccionado los ID en la entrada.

	A	B	C	D	E	F	G
	Text	Form	Topic Category	Rank	Type	Polarity Tag	Polarity Score
1	Las tarifas de algunos servicios básicos como la luz, el billete de tren o los precios de las autopistas vuelven a subir con el inicio del nuevo año.	autopista	concept	1	Top>Location>Facility>Line>Railroad	N	-0,6
2		servicio	concept	2	Top>Product>ProfessionalService	NONE	
3		tren	concept	3	Top>Product>Machine>Vehicle>Train	NONE	
4		precio	concept	4	Top>Numex>Money	NONE	
5		año	concept	5	Top>Time>Period	NONE	
6	El ministro de Economía, Luis de Guindos, ha asegurado que en 2014 habrá creación neta de empleo que será "incluso superior" a la proyectada en el mes de septiembre cuando se elaboraron los Presupuestos Generales del Estado.	De Guindos	entity	1	Top>Person>FullName	P	0,6
7		ministro de Economía	concept	1	Top	P	0,6
8		septiembre	concept	2	Top>Time>Date	P	0,6
9		mes	concept	3	Top>Time>Period	P	0,6
10		Presupuestos Generales del Estado	concept	4	Top	P	0,6
11		empleo	concept	5	Top>OtherEntity>Vocation	NONE	
12	Seis de las ocho mayores suspensiones de pagos del año las protagonizaron empresas vinculadas al negocio del ladrillo, como Fergo Aisa, Renta Corporación, Amrey y Contratas y Obras.	Fergo Aisa	entity	1	Top	NONE	
13		Renta Corporación	entity	2	Top	NONE	
14		año	concept	1	Top>Time>Period	N+	-1
15		compañía	concept	2	Top>Organization>Company	N+	-1
16	El Grupo Unidos por el Canal (GUPEC), el consorcio encargado de la ampliación del Canal de Panamá liderado por la española Sacyr, ha anunciado que suspenderá las obras debido a la imposibilidad de llevarlas a cabo con los costos pactados inicialmente en el contrato.	Canal	entity	1	Top>Person>LastName	P	0,6
17		Canal de Panamá	entity	2	Top	P	0,6
18		Grupo Unidos	entity	3	Top	P	0,6
19		Sacyr Vallehermoso	entity	4	Top>Organization>Company>IndustrialCompany>IndustrialSen	NEU	0
20		encargado	concept	1	Top>OtherEntity>Title>PositionTitle	P	0,6
21		asociación	concept	2	Top>Organization>Institute	P	0,6

**Ilustración 21: Ejemplo Topic Sentiment Analysis**

### 2.4.6 Otras

Existen muchas otras herramientas para el análisis automático de sentimientos como Skyttle, Textalytics, Text-processing, etc., pero se han elegido las anteriores para el estudio porque además de ser las más relevantes, tienen unas características que se adaptan mejor a las necesidades del proyecto.

## 2.5 Análisis y visualización de datos

### 2.5.1 Análisis

Después de la obtención de la puntuación de las noticias, vamos a realizar el análisis de los resultados.

### 2.5.2 Análisis de noticias

Porcentaje de noticias positivas, negativas y neutras. Hay un total de 28.630 noticias almacenadas en la base de datos clasificadas con la siguiente polaridad:

- Positivas: 13.931
- Negativas: 6.475
- Neutras: 3.714

None: 4.510

## 2.6 Metodología SCRUM

Scrum es el nombre con el que se denomina a los marcos de desarrollo ágiles caracterizados por:

- Adoptar una estrategia de desarrollo incremental, en lugar de la planificación y ejecución completa del producto.

## CAPÍTULO 2

- Basar la calidad del resultado más en el conocimiento tácito de las personas en equipos autoorganizados, que en la calidad de los procesos empleados.
- Solapamiento de las diferentes fases del desarrollo, en lugar de realizar una tras otra en un ciclo secuencial o de cascada.

SCRUM es un modelo de referencia que define un conjunto de prácticas y roles, y que puede tomarse como punto de partida para definir el proceso de desarrollo que se ejecutará durante un proyecto. Los roles principales en Scrum son el ScrumMaster, que procura facilitar la aplicación de scrum y gestionar cambios, el ProductOwner, que representa a los stakeholders (interesados externos o internos), y el Team que ejecuta el desarrollo y demás elementos relacionados con el. Durante cada sprint, un periodo entre una y cuatro semanas (la magnitud es definida por el equipo y debe ser lo mas corta posible), el equipo crea un incremento de software potencialmente entregable (utilizable). El conjunto de características que forma parte de cada sprint viene del Product Backlog, que es un conjunto de requisitos de alto nivel priorizados que definen el trabajo a realizar (PBI, Product Backlog Item). Los elementos del Product Backlog que forman parte del sprint se determinan durante la reunión de Sprint Planning. Durante esta reunión, el Product Owner identifica los elementos del Product Backlog que quiere ver completados y los hace del conocimiento del equipo. Entonces, el equipo conversa con el Product Owner buscando claridad y magnitud adecuadas (Cumpliendo el INVEST) para luego determinar la cantidad de ese trabajo que puede comprometerse a completar durante el siguiente sprint.<sup>1</sup> Durante el sprint, nadie puede cambiar el Sprint Backlog, lo que significa que los requisitos están congelados durante el sprint.

Scrum permite la creación de equipos autoorganizados impulsando la co-localización de todos los miembros del equipo, y la comunicación verbal entre todos los miembros y disciplinas involucrados en el proyecto.

Un principio clave de Scrum es el reconocimiento de que durante un proyecto los clientes pueden cambiar de idea sobre lo que quieren y necesitan (a menudo llamado requirements churn), y que los desafíos impredecibles no pueden ser fácilmente enfrentados de una forma predictiva y planificada. Por lo tanto, Scrum adopta una aproximación pragmática, aceptando que el problema no puede ser completamente entendido o definido, y centrándose en maximizar la capacidad del equipo de entregar rápidamente y responder a requisitos emergentes.

Las características más marcadas que se logran notar en Scrum serían: gestión regular de las expectativas del cliente, resultados anticipados, flexibilidad y adaptación, retorno de inversión, mitigación de riesgos, productividad y calidad, alineamiento entre cliente y equipo, por último equipo motivado. Cada uno de estos puntos mencionados hacen que el Scrum sea utilizado de manera regular en un conjunto de buenas prácticas para el trabajo en equipo y de esa manera obtener resultados posibles.

Existen varias implementaciones de sistemas para gestionar el proceso de Scrum, que van desde notas amarillas "post-it" y pizarras hasta paquetes de software. Una de las mayores ventajas de Scrum es que es muy fácil de aprender, y requiere muy poco esfuerzo para comenzarse a utilizar.



# Capítulo 6

## Análisis, diseño e implementación del sistema

### 3.1 Análisis de requisitos de usuario

En este apartado se va a realizar un análisis del sistema mediante los requisitos de usuario. Estos requisitos se encargarán de dar información acerca de las funcionalidades del sistema y de sus restricciones. Dado que no es el objetivo principal del Proyecto de Fin de Carrera se expondrán de manera informal.

Un requisito es una “condición o capacidad que necesita el usuario para resolver un problema o conseguir un objetivo determinado”. (Landazabal, 2008)

Los requisitos de usuario están divididos en requisitos funcionales y requisitos no funcionales.

Los requisitos funcionales expresan la naturaleza del sistema (como interactúa el sistema con su entorno y cuál va a ser su estado y funcionamiento), los servicios o funciones que proveerá el sistema y describen la interacción entre el sistema y el entorno.

Los requisitos no funcionales son restricciones a los servicios o funciones ofrecidos por el sistema. Describen restricciones que limitan las elecciones para construir una solución.

### 3.1.1 Requisitos funcionales

Los requisitos funcionales definidos para este sistema son los siguientes:

Un requisito principal de este proyecto es la obtención de los valores de cierre diario de las distintas empresas cotizadas en la Bolsa de Madrid. Éstos pueden ser obtenidos mediante la web principal de la Bolsa de Madrid <http://www.bolsamadrid.es/>. Su prioridad es alta.

Otro requisito principal del proyecto es la obtención de noticias relacionadas con la economía. Para ello se decide obtener la información de la hemeroteca online del periódico Expansión. Éstos pueden ser obtenidos mediante en la siguiente web <http://www.expansion.com/hemeroteca/2014/>. Su prioridad es alta.

Una vez que tenemos los valores de la Bolsa de Madrid y las noticias es necesario realizar el análisis de sentimiento de las noticias con la herramienta MeaningCloud. Su prioridad es alta.

Con toda la información necesaria para realizar el análisis, necesitamos almacenar esta información en una base de datos. En este caso se almacenarán en Access 2010. Su prioridad es media.

### 3.1.2 Requisitos no funcionales

Como requisitos no funcionales se definen en esta memoria la comparación de las diferentes herramientas de análisis de sentimiento con una descripción, costes y limitaciones de cada una. Su prioridad es baja.

Se hará un análisis previo de trabajos similares. Su prioridad es baja.

## 3.1 Diseño e implementación

La aplicación “Analizador Bolsa” se ha desarrollado en VBA en una base de datos Access 2010.

Esta aplicación consta de una pantalla principal con los siguientes elementos:

- Combobox **empresas**. Las empresas cotizantes en la Bolsa de Madrid. Este combo se carga inicialmente de la tabla BME.
- Combobox **fecha**. Inicialmente vacío. Cuando se selecciona una empresa, se cargan los días en los que hay noticias que hacen referencia a la misma. Cuando se selecciona una noticia, se selecciona la fecha en la que fue publicada.

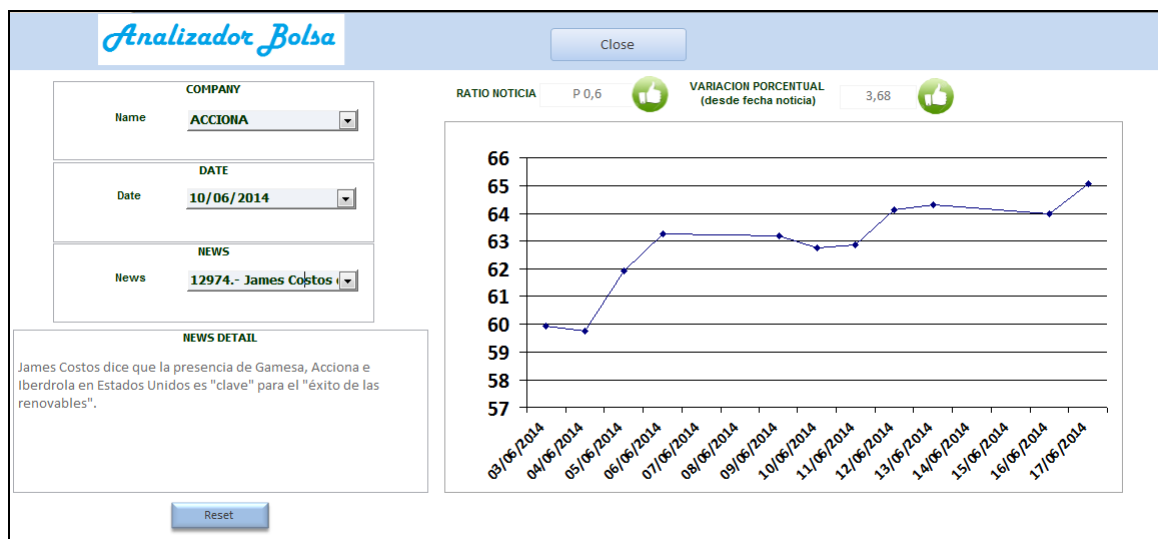
- Combobox **noticias**. Inicialmente vacío. Cuando se selecciona una empresa, se cargan las noticias que hacen referencia a la misma. Cuando se selecciona una fecha, se cargan las noticias de la empresa seleccionada en la fecha seleccionada.
- Textbox **noticia**. Cuando se selecciona una noticia, se carga la noticia completa para poder leerla completamente.
- Botón **reset**. Inicializa la aplicación a la situación inicial, para comenzar una nueva búsqueda.

The screenshot shows the initial interface of the 'Analizador Bolsa' application. It features a light blue header bar with the application's name 'Analizador Bolsa' on the left and a 'Close' button on the right. The main content area is divided into several sections. The first section is titled 'COMPANY' and contains a 'Name' dropdown menu. The second section is titled 'DATE' and contains a 'Date' dropdown menu. The third section is titled 'NEWS' and contains a 'News' dropdown menu. Below these three sections is a large, empty rectangular area labeled 'NEWS DETAIL'. At the bottom center of the interface is a 'Reset' button.

**Ilustración 22: Pantalla inicial “Analizador Bolsa”**

Al seleccionar una noticia, además de completarse el textbox **noticia**, se muestra un gráfico y dos valores que se explican a continuación:

- **Ratio noticia**. Puntuación de la noticia calculada anteriormente con el analizador de sentimientos Meaning Cloud. Valor entre -1 y 1.
- **Valoración porcentual**. Cálculo de la variación desde el día de la noticia y 7 días después.
- **Gráfico**. Visualización del gráfico que ilustra la variación del cierre diario en bolsa para la empresa seleccionada en las fechas comprendidas entre los 7 días anteriores y 7 días posteriores a la fecha de la noticia.



**Ilustración 23: Gráfico de la aplicación “Analizador Bolsa”**

En la visualización del gráfico y los valores del Ratio de la noticia y de la variación porcentual se pueden sacar conclusiones sobre el comportamiento de la bolsa a raíz de una noticia.

# Capítulo 7

## Pruebas

Para comprobar el funcionamiento de la herramienta se han realizado una serie de pruebas tanto funcionales como no funcionales. Para cada prueba se han ido cambiando tanto las noticias como los parámetros adaptándose a las entradas y las salidas.

### 4.1 Pruebas funcionales

“Una prueba funcional es una prueba basada en la ejecución, revisión y retroalimentación de las funcionalidades previamente diseñadas para el software. Las pruebas funcionales se hacen mediante el diseño de modelos de prueba que buscan evaluar cada una de las opciones con las que cuenta el paquete informático. Dicho de otro modo son pruebas específicas, concretas y exhaustivas para probar y validar que el software hace lo que debe y sobre todo, lo que se ha especificado.” ([Definición Wikipedia](#))

A continuación se muestran los resultados de las pruebas funcionales realizadas sobre la herramienta “Analizador Bolsa”.

#### 1. Combobox ‘Company’

Al seleccionar una empresa en el combobox ‘Company’ se lanza el evento *Company\_Change()*. Este evento realiza las siguientes acciones:

- Si los combobox ‘Date’ y/o ‘News’ contienen datos (inicialmente se cargan vacíos) se vacían.

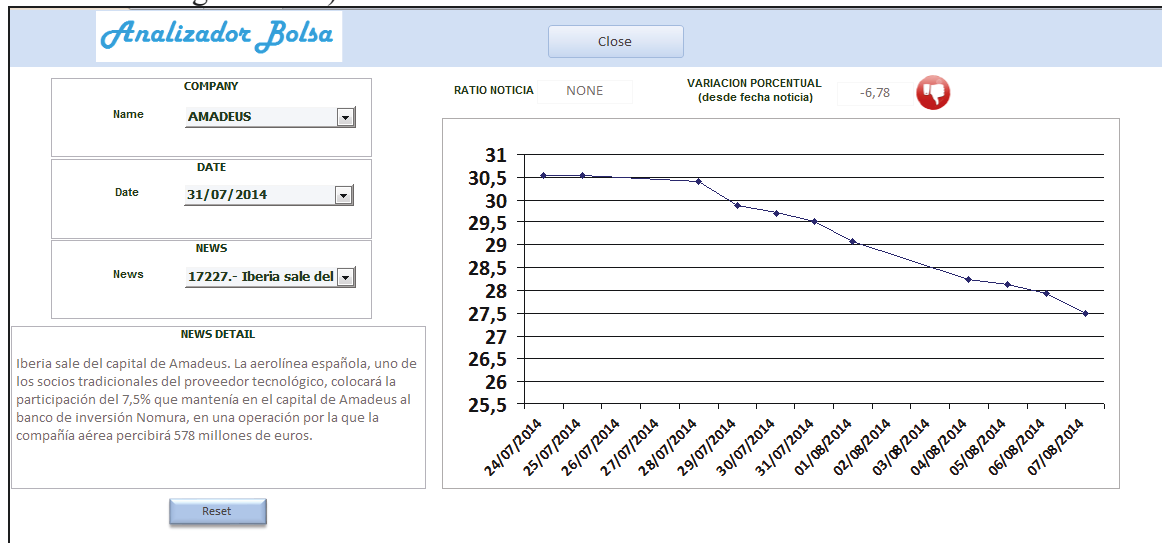


Ilustración 24: Prueba 1.1

**Analizador Bolsa**

Close

COMPANY  
Name: BBVA

DATE  
Date:

NEWS  
News:

NEWS DETAIL

Reset

Ilustración 25: Prueba 1.2

- Hay dos opciones, que se seleccione el ítem ‘vacío’ o que se seleccione una empresa. En este caso se re-cargan las fechas y noticias relacionadas con la empresa. Además se vacía la caja de texto ‘News Details’ y se ocultan los ítems que componen el resultado (Ratio noticia, Variación porcentual y el gráfico).

The screenshot shows the 'Analizador Bolsa' application window. At the top, there is a title bar with the application name and a 'Close' button. Below the title bar, there are three main sections: 'COMPANY', 'DATE', and 'NEWS'. The 'COMPANY' section has a 'Name' dropdown menu with 'BBVA' selected. The 'DATE' section has a 'Date' dropdown menu that is open, displaying a list of dates from 07/01/2014 to 30/01/2014. The 'NEWS' section has a 'News' dropdown menu that is also open, displaying a list of news items. A 'Reset' button is located at the bottom of the window.

Ilustración 26: Prueba 1.3

The screenshot shows the 'Analizador Bolsa' application window. At the top, there is a title bar with the application name and a 'Close' button. Below the title bar, there are three main sections: 'COMPANY', 'DATE', and 'NEWS'. The 'COMPANY' section has a 'Name' dropdown menu with 'BBVA' selected. The 'DATE' section has a 'Date' dropdown menu that is open, displaying a list of dates from 07/01/2014 to 30/01/2014. The 'NEWS' section has a 'News' dropdown menu that is also open, displaying a list of news items. A 'Reset' button is located at the bottom of the window.

Ilustración 27: Prueba 1.4

## 2. Combobox 'Date'

Al seleccionar una fecha en el combobox 'Date' se lanza el evento *DateCombo\_Change()*. Este evento realiza las siguientes acciones:

- Si en el combobox 'Date' se ha seleccionado una fecha (no vacío) se recargan las noticias en el combobox 'News' relacionadas con la empresa y la fecha seleccionada.

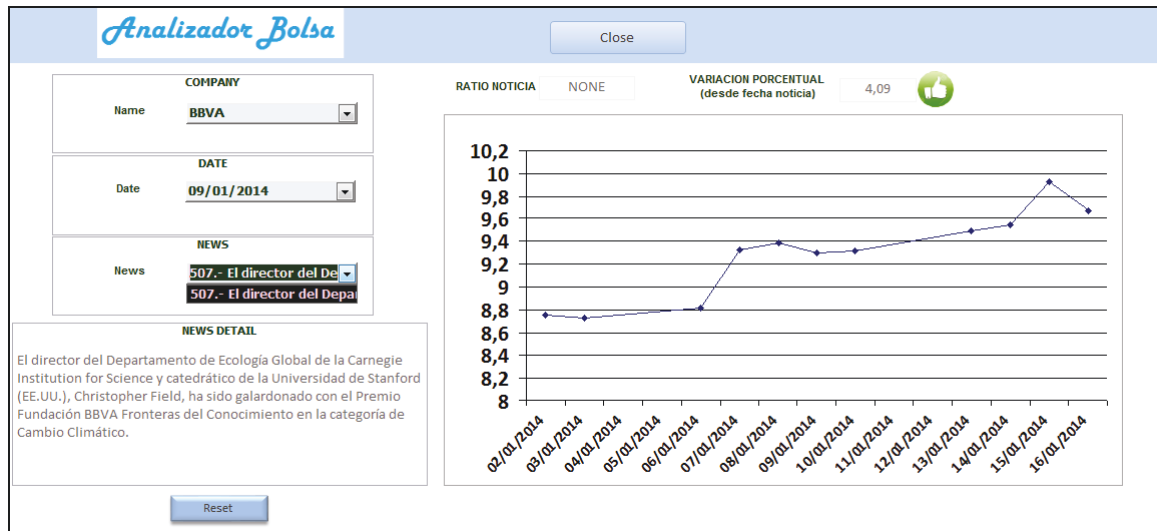


Ilustración 28: Prueba 2.1

**Analizador Bolsa**

Close

**COMPANY**

Name: **BBVA**

**DATE**

Date: **14/01/2014**

**NEWS**

News: **1240.- Marvin Minsky, cor**

**NEWS DETAIL**

Reset

Ilustración 29: Prueba 2.2

- Si en el combobox 'Date' se ha seleccionado el ítem 'vacío' se recargan todas las noticias en el combobox 'News' relacionadas con la empresa seleccionada.



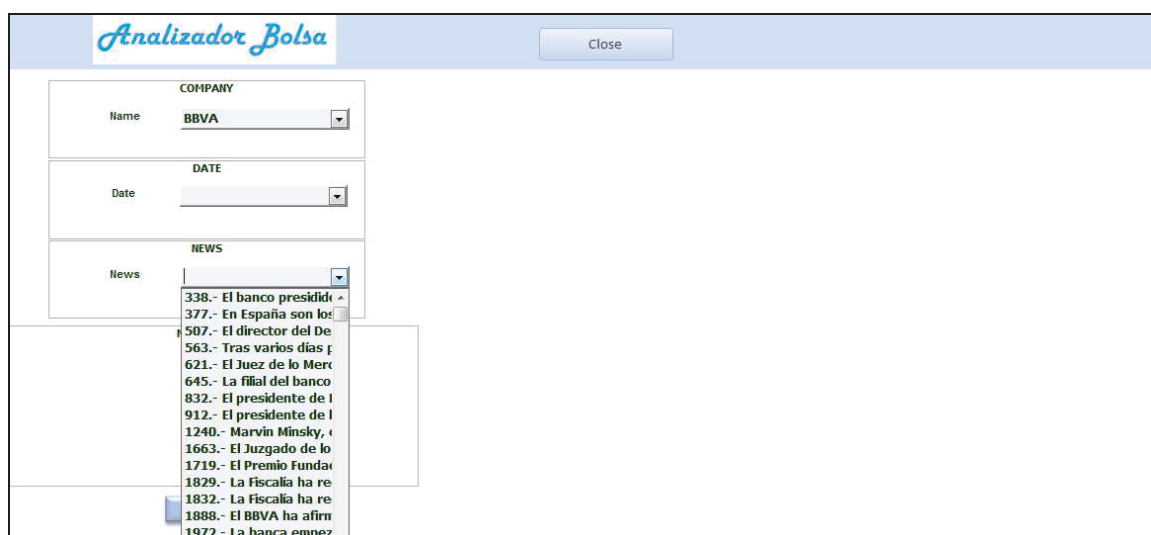




Ilustración 30: Prueba 2.3

### 3. Combobox 'News'

Al seleccionar una noticia en el combobox 'News' se lanza el evento *News\_Change()*. Este evento realiza las siguientes acciones:

- Si en el combobox 'Date' se ha seleccionado una fecha (no vacío) se recargan las noticias en el combobox 'News' relacionadas con la empresa y la fecha seleccionada.
- Se pone visible el textbox 'Ratio Noticia' junto al icono positivo o

negativo. Si el Ratio de la noticia es positivo se muestra el icono  y

si el Ratio es negativo se muestra el icono . Se realiza lo mismo con el textbox 'Variación Pocental' con la misma lógica en la visualización de los iconos. Se lanza también una consulta a la base de datos recuperando la Fecha y Cierre de la tabla BME donde la fecha sea mayor o igual a 7 días atrás de la fecha de la noticia y menor o igual a 7 días posteriores a la fecha de la noticia (por ejemplo, si la fecha de la noticia es el 14 de Agosto, se recuperan los cierres desde el 7 de Agosto incluido hasta el 21 de Agosto incluido).

Se calcula el Valor Porcentual de la siguiente manera:

$$VP = ((V1 - V2) / V1) * 100$$

Donde VP es el Valor Porcentual, V1 es el valor de cierre de la fecha de la noticia o si ese día es festivo o fin de semana se recupera el día anterior. V2 es el valor del cierre 7 días después de la noticia.

### 4. Botón 'Reset'

## CAPÍTULO 7

Al pulsar en el botón ‘RESET’ de la herramienta, la aplicación realiza las siguientes acciones:

- Los combobox ‘Company’, ‘Date’ y ‘New’ se inicializan a vacío.
- El gráfico se oculta
- Se añade un ítem vacío en el combobox ‘Company’
- Se cargan todas las empresas almacenadas en la tabla ‘BME’ en el combobox ‘Company’

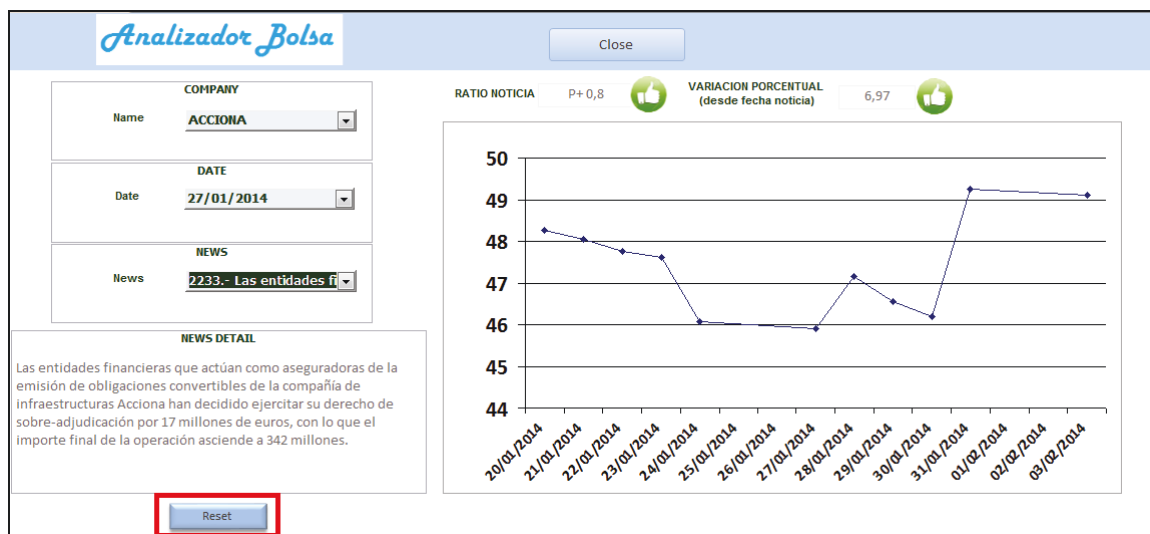


Ilustración 31: Prueba 4.1

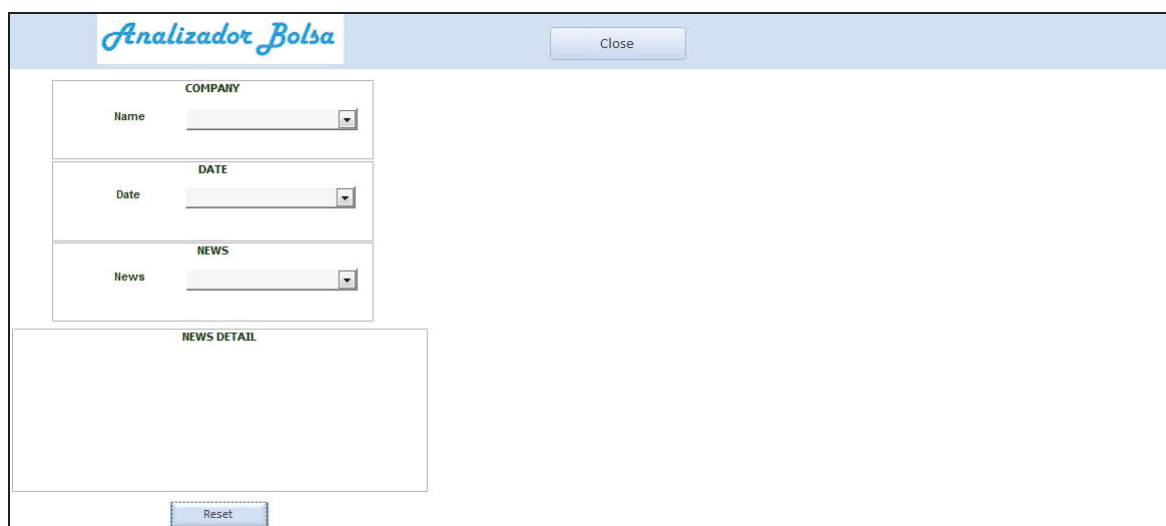


Ilustración 32: Prueba 4.2

## 4.2 Pruebas no funcionales

Las pruebas no funcionales son aquellas que no prueban el funcionamiento de los componentes de la herramienta:

- Volumen de pruebas. Las pruebas de volumetría miden la capacidad de la aplicación tanto para un alto volumen de entrada y salida de datos de la base de datos como de los campos existentes en la herramienta.
- Prueba de carga. Las pruebas de carga o esfuerzo son aquellas que tienen como finalidad analizar la respuesta de la herramienta con cargas masivas de datos, repeticiones de acciones.
- Pruebas de rendimiento. Las pruebas de rendimiento se realizan para medir tiempos de respuesta.
- Pruebas de configuración. Pruebas para verificar si la herramienta puede ejecutarse en diferentes versiones o configuraciones de software y hardware.

# Capítulo 8

## Planificación

Este apartado está dividido en dos secciones: en la primera se muestra una planificación que se hizo inicialmente de manera orientativa y en la segunda se definirá la ejecución del proyecto que se ha realizado en realidad.

Las variaciones se deben a ciertos problemas técnicos que han surgido durante la elaboración del proyecto y a nuevas ideas que han ido surgiendo y se han ido adaptando a lo largo de este.

Para ambas planificaciones se ha utilizado el diagrama de Gantt. Esta es una herramienta gráfica cuyo objetivo es exponer el tiempo de dedicación previsto para diferentes tareas o actividades a lo largo de un tiempo total determinado. Es importante tener en cuenta que el diagrama de Gantt no indica las relaciones existentes entre actividades, solo el tiempo invertido en cada una de ellas.

Este diagrama ha sido creado con la herramienta MS Project. Es un software de administración de proyectos diseñado, desarrollado y comercializado por Microsoft para asistir a administradores de proyectos en el desarrollo de planes, asignación de recursos a tareas, dar seguimiento al progreso, administrar presupuesto y analizar cargas de trabajo.

En los gráficos obtenidos con el MS Project no se puede obtener una visión real del tiempo invertido en cada fase, ya que por motivos laborales no puedo invertir el mismo número de horas cada día. Por ello también se realiza un gráfico con las horas invertidas en cada fase.

La planificación se ha realizado en base a las fases necesarias para el ciclo de vida de un proyecto (Fernando Berzal, El ciclo de vida de un sistema de información). Estas fases son:

1. **Planificación:** Antes de iniciar un proyecto es necesario realizar una serie de tareas previas que influirán decisivamente en la finalización con éxito del proyecto. Estas tareas se conocen popularmente como el *fuzzy front-end* del proyecto al no estar sujetas a plazos. Las tareas iniciales que se realizarán en esta fase inicial del proyecto incluyen actividades tales como la determinación del ámbito del proyecto, la realización de un estudio de viabilidad, una estimación del coste del proyecto, su planificación temporal y la asignación de recursos a las distintas etapas del proyecto.
2. **Análisis:** Lo primero que se debe hacer cuando quieres construir un sistema de información es averiguar qué es exactamente lo que tiene que hacer el sistema. La etapa de análisis en el ciclo de vida del software corresponde al proceso mediante el cual se intenta descubrir qué es lo que realmente se necesita y se llega a una comprensión adecuada de las características que el sistema debe poseer. En esta fase se deben analizar los requerimientos del programa y se debe analizar la situación actual del ámbito del proyecto.
3. **Diseño:** Mientras que los modelos utilizados en la etapa de análisis representan los requisitos del usuario desde distintos puntos de vista (el qué), los modelos que se utilizan en la fase de diseño representan las características del sistema que nos permitirán implementarlo de forma efectiva (el cómo).
4. **Implementación:** Una vez que sabemos qué funciones debe desempeñar nuestro sistema de información (análisis) y hemos decidido cómo vamos a organizar sus distintos componentes (diseño), es el momento de pasar a la etapa de implementación.
5. **Pruebas:** la etapa de pruebas tiene como objetivo detectar los errores que se hayan podido cometer en las etapas anteriores del proyecto.

Además de estas fases se realiza una tarea de documentación a lo largo de todo el proyecto.

En cada una de las fases para las dos planificaciones se detallará el número de horas invertidas y las personas que han participado.

## 5.1 Ciclo de vida iterativo

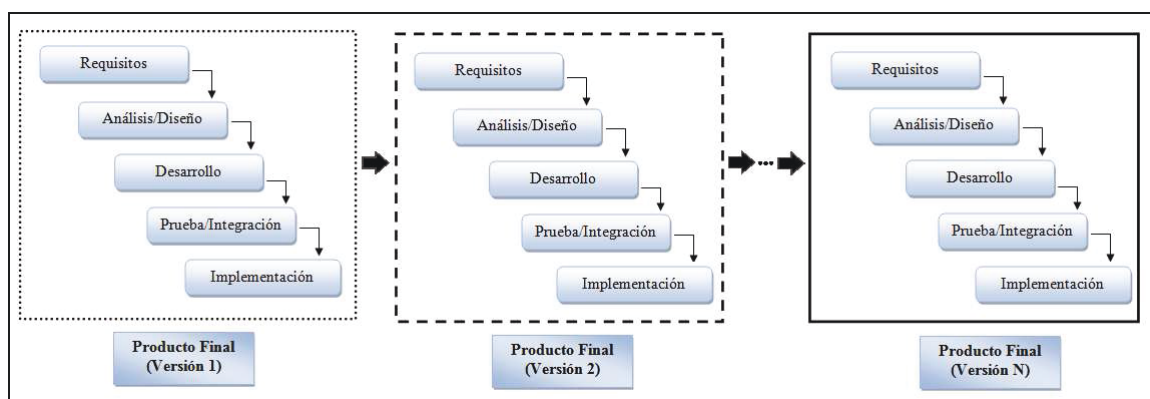
Este proyecto se ha desarrollado en base al modelo del ciclo de vida iterativo principalmente por las siguientes razones:

- Mejora las debilidades del modelo tradicional de cascada

- Provee de soporte para determinar la efectividad de los procesos y de la calidad del producto
- Permite estudiar y después mejorar y ajustar el proceso para el ambiente en particular

El modelo consta de varias etapas de desarrollo en cada incremento, las cuales inician con el análisis y finalizan con la instauración y aprobación del sistema.

Lo que se busca es que en cada iteración los componentes logren evolucionar el producto dependiendo de los completados de las iteraciones antecesoras, agregando más opciones de requisitos y logrando así un mejoramiento mucho más completo.



**Ilustración 33: Esque del Ciclo de Vida Iterativo**

En primer lugar se realizó una planificación del sistema, con las fechas estimadas en las que se iba a realizar cada tarea, un presupuesto inicial, y una visión inicial de la viabilidad del proyecto.

En el análisis se realizó un proceso de recopilación de los requisitos tanto funcionales como no funcionales del sistema.

Después se diseñó el sistema centrándose en la estructura de los datos, la arquitectura del software y el detalle procedimental. En esta etapa se hizo una representación gráfica del sistema a partir de los requisitos obtenidos para facilitar la codificación.

Una vez que se realizó el diseño tuvo lugar la implementación.

Cuando se había generado el código se realizaron las pruebas oportunas para verificar el correcto funcionamiento del sistema.

Y una vez que se realizó la aplicación, nos encontramos en la fase de mantenimiento y mejora continua.

En todas estas fases ha habido una labor de documentación continua.

## 5.2 Planificación inicial

En esta planificación inicial se indican las fases necesarias para el ciclo de vida del proyecto. La planificación inicial estaba definida con una duración de 395 horas.

### 1. Planificación.

#### a. Determinación del ámbito del proyecto:

- Tiempo estimado: 8 horas.
- Participantes: Harith Taha Abdulla Aljumaily, Dolores Cuadra y Ismael Rodrigo Huete.

#### b. Realización de un estudio de viabilidad:

- Tiempo estimado: 10 horas.
- Participantes: Harith Taha Abdulla Aljumaily, Dolores Cuadra y Ismael Rodrigo Huete.

#### c. Estimación del coste del proyecto:

- Tiempo estimado: 4 horas.
- Participantes: Ismael Rodrigo Huete.

#### d. Planificación temporal:

- Tiempo estimado: 4 horas.
- Participantes: Ismael Rodrigo Huete.

#### e. Asignación de recursos:

- Tiempo estimado: 2 horas.
- Participantes: Ismael Rodrigo Huete.

El total de horas estimadas para la planificación es: 28 horas.

### 2. Análisis.

#### a. Análisis del estado del arte:

- Tiempo estimado: 92 horas.
- Participantes: Ismael Rodrigo Huete.

#### b. Análisis de los requerimientos del programa:

- Tiempo estimado: 24 horas.
- Participantes: Harith Taha Abdulla Aljumaily, Dolores Cuadra y Ismael Rodrigo Huete.

El total de horas estimadas para el análisis es: 116 horas.

El total de horas estimadas para el análisis es: 116 horas.

### 3. **Diseño.**

#### a. Definición de la arquitectura:

- Tiempo estimado: 12 horas.
- Participantes: Ismael Rodrigo Huete.

#### b. Diseño de la interoperabilidad de las herramientas:

- Tiempo estimado: 14 horas.
- Participantes: Ismael Rodrigo Huete.

El total de horas estimadas para el diseño es: 26 horas.

### 4. **Implementación.**

#### a. Desarrollo de los clientes para las APIs utilizadas:

- Tiempo estimado: 60 horas.
- Participantes: Ismael Rodrigo Huete.

#### b. Desarrollo de los métodos de análisis:

- Tiempo estimado: 30 horas.
- Participantes: Ismael Rodrigo Huete.

#### c. Desarrollo de los métodos de clasificación:

- Tiempo estimado: 10 horas.
- Participantes: Ismael Rodrigo Huete.

#### d. Desarrollo de los algoritmos de validación:

- Tiempo estimado: 10 horas.
- Participantes: Ismael Rodrigo Huete.

El total de horas estimadas para la implementación es: 110 horas.

### 5. **Pruebas.**

#### a. Pruebas unitarias de cada herramienta instalada:

- Tiempo estimado: 22 horas.
- Participantes: Ismael Rodrigo Huete.



## b. Pruebas globales:

- Tiempo estimado: 20 horas.
- Participantes: Ismael Rodrigo Huete.

El total de horas estimadas para las pruebas es: 42 horas.

6. **Documentación.**

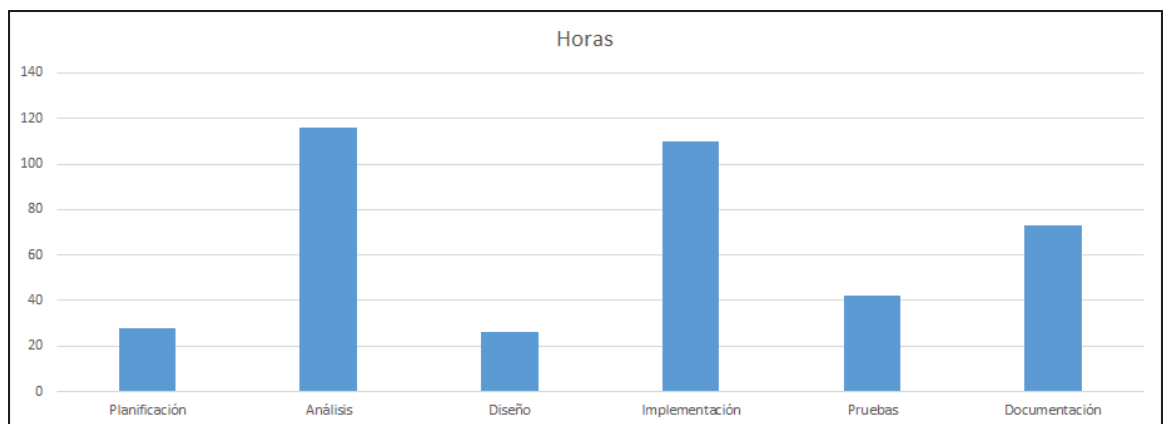
## a. Documentación del proyecto:

- Tiempo estimado: 65 horas.
- Participantes:

## b. Presentación:

- Tiempo estimado: 8 horas.
- Participantes:

El total de horas estimadas para la documentación es de: 73 horas.



**Ilustración 34: Costes de equipamiento**

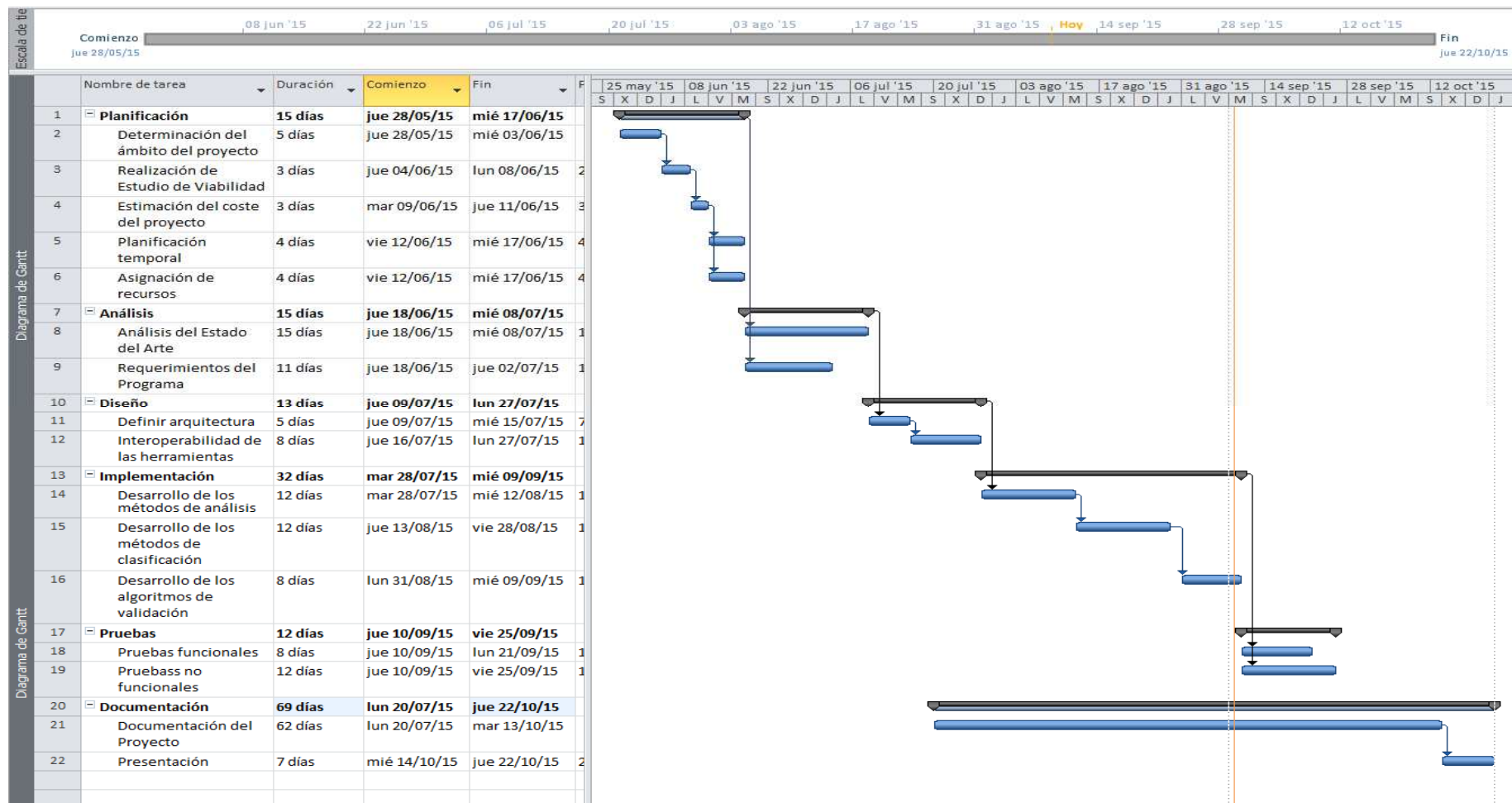


Ilustración 35: Diagrama de Gantt. Planificación estimada

## 5.3 Ejecución final del proyecto

En esta sección se indican tanto las fases necesarias para el ciclo de vida del proyecto como la descripción detallada de cada una de ellas. Además se comentará por qué se ha producido el desfase de las horas iniciales.

### 1. Planificación.

- a. Determinación del ámbito del proyecto: En varias reuniones mantenidas con los tutores del proyecto en las que me explicó una gran diversidad de acciones que se podían realizar con el tema elegido determinamos que el ámbito del proyecto se iba a centrar en analizar el comportamiento de la bolsa a partir de las noticias.
  - Tiempo estimado: 8 horas.
  - Participantes: Harith Taha Abdulla Aljumaily, Dolores Cuadra y Ismael Rodrigo Huete.
- b. Realización de un estudio de viabilidad: A raíz de esas reuniones se buscó información para estudiar la viabilidad del proyecto.
  - Tiempo estimado: 10 horas.
  - Participantes: Harith Taha Abdulla Aljumaily, Dolores Cuadra y Ismael Rodrigo Huete.
- c. Estimación del coste del proyecto: Se hizo una estimación inicial teniendo en cuenta unas herramientas iniciales que han tenido que ser modificadas y con los recursos disponibles.
  - Tiempo estimado: 4 horas.
  - Participantes: Ismael Rodrigo Huete.
- d. Planificación temporal: De manera orientativa también se realizó la planificación que se ha descrito anteriormente.
  - Tiempo estimado: 4 horas.
  - Participantes: Ismael Rodrigo Huete.
- e. Asignación de recursos: Se asignaron los recursos a todas las fases del proyecto.
  - Tiempo estimado: 2 horas.
  - Participantes: Ismael Rodrigo Huete.

## CAPÍTULO 8

El total de horas estimadas que duró la planificación no supuso ningún cambio en relación con lo que estaba estimado. Esta duración fue de 28 horas.

### 2. Análisis.

- a. Análisis del estado del arte: Antes de crear el motor era necesario tener una visión actual de las herramientas que existían ya en el mercado para la realización de temas parecidos a los del proyecto y conseguir la información necesaria para el estudio. Ésta fue una gran tarea de investigación y documentación que se ha desarrollado a lo largo de todo el proyecto.

- Tiempo estimado: 92 horas.
- Participantes: Ismael Rodrigo Huete.

- b. Análisis de los requerimientos del programa: Para la realización del proyecto era necesario utilizar una serie de herramientas y cada una de ellas necesitaba ser analizada.

- Tiempo estimado: 24 horas.
- Participantes: Harith Taha Abdulla Aljumaily, Dolores Cuadra y Ismael Rodrigo Huete.

El total de horas invertidas en el análisis es: 116 horas.

### 3. Diseño.

- a. Definición de la arquitectura: En esta fase se diseña la arquitectura del proyecto.

- Tiempo estimado: 12 horas.
- Participantes: Ismael Rodrigo Huete.

- b. Diseño de la interoperabilidad de las herramientas: Como se van a utilizar distintas herramientas en el proyecto se diseña la manera en la que van a interactuar.

- Tiempo estimado: 14 horas.
- Participantes: Ismael Rodrigo Huete.

El total de horas invertidas en el diseño es: 26 horas.

### 4. Implementación.

- a. Desarrollo y extracción de sentimientos de las noticias: Se realiza el proyecto en VBA con Access 2010. El punto más crítico de todo el proyecto es el siguiente: en el diseño y el análisis se propone realizar el análisis de sentimientos de las noticias con herramientas con un periodo de prueba. Pero muchas de las herramientas que se iban a utilizar no funcionaban correctamente, o no tienen la funcionalidad en

español como venía en su descripción por lo que es necesario cambiar de herramientas y volver a estudiar el funcionamiento de cada una de ellas. Finalmente se utiliza el add-in de Excel Meaning Cloud. Esto hace que se retrase el proyecto un largo periodo de tiempo.

- Tiempo estimado: 80 horas.
- Participantes: Ismael Rodrigo Huete.

b. Desarrollo de los métodos de análisis: Los métodos de análisis se realizan a la par que el desarrollo de los clientes por lo que también sufre variación.

- Tiempo estimado: 45 horas.
- Participantes: Ismael Rodrigo Huete.

c. Desarrollo de los métodos de clasificación: Se limitan los métodos de clasificación a partir de los que había diseñados. Siguen realizando la misma función pero no se obtiene tanto detalle.

- Tiempo estimado: 12 horas.
- Participantes: Ismael Rodrigo Huete.

d. Desarrollo de los algoritmos de validación: Los algoritmos de validación siguen la planificación inicial.

- Tiempo estimado: 12 horas.
- Participantes: Ismael Rodrigo Huete.

El total de horas invertidas en la implementación es: 149 horas.

## 5. Pruebas.

a. Pruebas unitarias de cada herramienta instalada: Únicamente se realizan pruebas con las herramientas válidas por lo que tampoco difiere en exceso de la planificación inicial.

- Tiempo estimado: 22 horas.
- Participantes: Ismael Rodrigo Huete.

b. Pruebas globales: Se realizan más pruebas de las que se habían planificado en un principio.

- Tiempo estimado: 24 horas.
- Participantes: Ismael Rodrigo Huete.

El total de horas invertidas en las pruebas es: 46 horas.

## 6. Documentación.

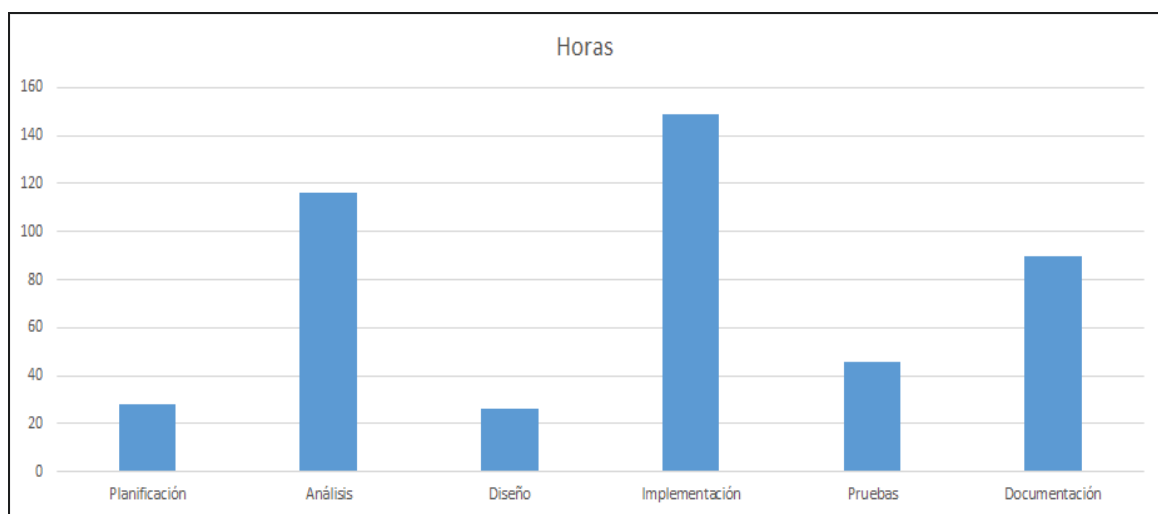
## CAPÍTULO 8

- a. Documentación del proyecto: La documentación del proyecto se realiza a lo largo de todo el proyecto junto con el análisis del estado del arte, están fuertemente conectados el uno con el otro.
  - Tiempo estimado: 78 horas.
  - Participantes: Ismael Rodrigo Huete.
- b. Presentación: Una vez terminada tanto la implementación como el documento se realiza una presentación con un resumen del proyecto.
  - Tiempo estimado: 12 horas.
  - Participantes: Ismael Rodrigo Huete.

El total de horas invertidas en la documentación es de: 90 horas.

La duración real del proyecto es de **445 horas**.

Como he comentado anteriormente tanto la planificación inicial como la planificación real está definida en horas en vez de días ya que no son 8 exactamente las horas invertidas en el trabajo, si no que esta cifra fluctúa dependiendo del día por motivos laborales. Por este motivo veo más efectivo realizar la estimación en horas.



**Ilustración 36: Costes de equipamiento**

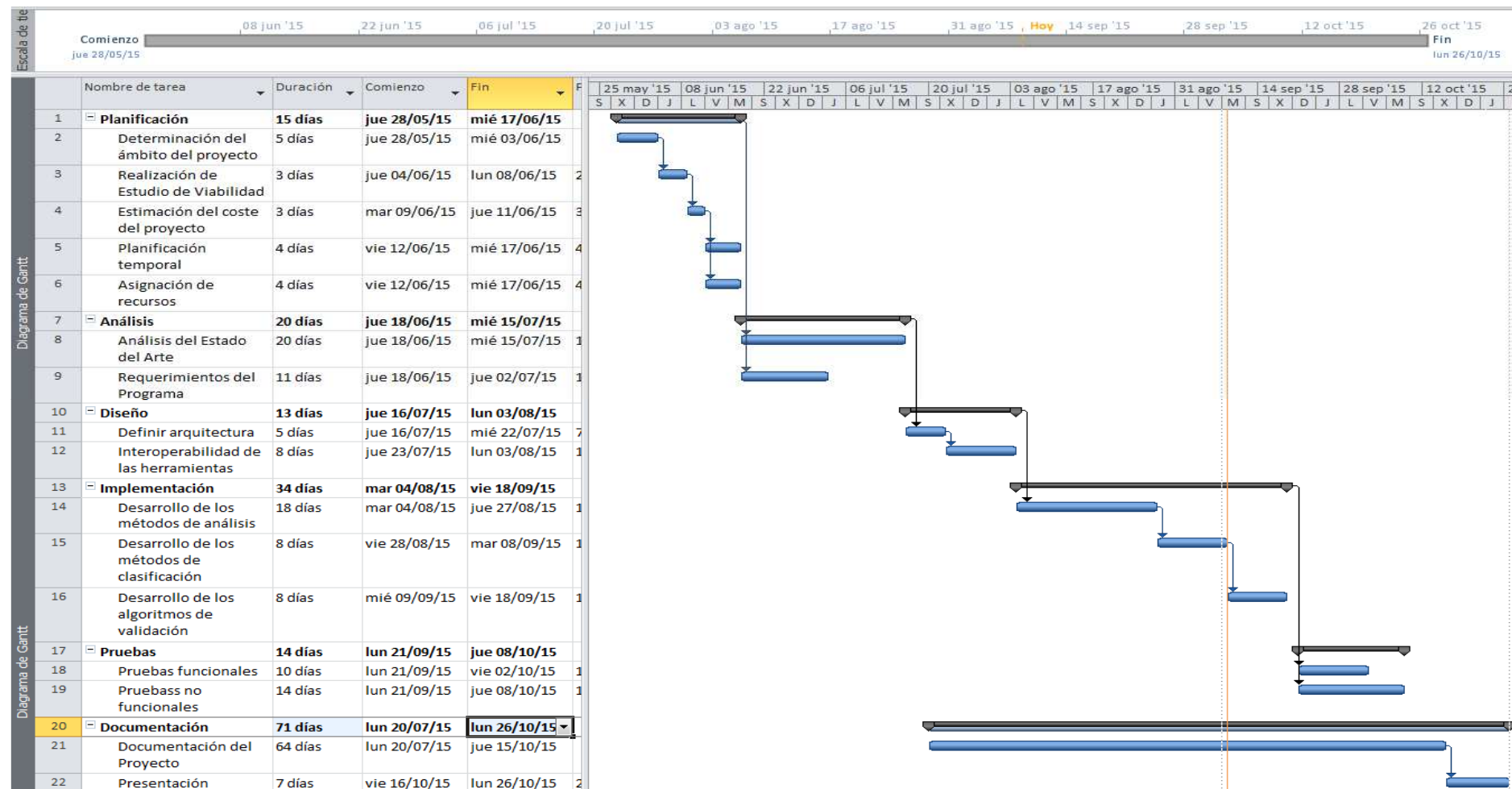


Ilustración 37: Diagrama de Gantt. Planificación final

# Capítulo 9

## Presupuesto

En el presupuesto se tienen en cuenta tanto el número de horas invertidas por las personas que realizan el proyecto como las herramientas utilizadas para el funcionamiento de este.

Primeramente, como se ha visto en la fase de planificación, se crea un presupuesto inicial y una vez finalizado el proyecto se hace una valoración económica real.

Para valorar el coste que supone cada una de las personas que realizan el proyecto, además de tener en cuenta su tarifa, es necesario prever los costes de la seguridad social, retenciones, etc.

La persona que realiza el proyecto tiene que ser un Ingeniero Técnico Informático Junior con conocimiento en programación VBA. Además tiene que tener la capacidad de gestionar el proyecto ya que va a estar inmerso en él desde la planificación hasta la documentación. No es necesario el conocimiento de las herramientas que se van a utilizar aunque sí es recomendable, pero ese conocimiento se irá adquiriendo a lo largo del proyecto.

La persona encargada de tutelar el proyecto tiene que ser Ingeniero Informático Senior, profesor de la Universidad Carlos III y tener experiencia en la gestión de proyectos de fin de grado.



Se adjunta como anexo (Anexo I. Cálculo Salario Profesionales) el cálculo del salario neto de los profesionales asignados al proyecto para comprobar su viabilidad y el cálculo del coste/hora de cada uno.

Teniendo en cuenta la información recogida en el anexo, el coste/hora de cada profesional es el siguiente:

- Ingeniero Técnico Jr.: 11,06 €/h
- Ingeniero Sr.: 22,12 €/h

Ambos presupuestos están estimados a partir de las horas de trabajo realizadas. Se elige realizar la estimación en horas en vez de días porque por motivos laborales no se puede invertir 8 horas diarias en el proyecto.

El coste material vendrá definido por el coste de las herramientas tanto hardware como software necesarias para realizar el proyecto.

Los costes relacionados con las dietas, traslados, costes de estructura, etc. serán asumidos por las personas que realizan el proyecto por lo que no supondrán ningún coste adicional.

## 6.1 Presupuesto inicial

En esta sección se incluyen los costes totales del proyecto descompuestos en costes de personal y de equipamiento. Para evitar confusiones se debe comentar que todas las cantidades están expresadas en euros.

### 6.1.1 Costes de Personal

Para calcular el coste del personal se han tenido en cuenta las dos personas relacionadas con el proyecto, la función de cada una de ellas y el número de horas invertidas.

A continuación se encuentra desglosado el coste personal del proyecto según las diferentes actividades realizadas y previamente descritas. Todos estos costes no incluyen IVA.

	Perfil	Tarifa/ Hora	Horas Totales	Coste Total
Recurso 1	Ingeniero Téc. Informático Jr.	11,06	360	3.981,60
Recurso 2	Ingeniero Informático Sr.	22,12	35	774,20
<b>Coste Total Personal</b>			<b>395</b>	<b>4.755,08</b>

**Ilustración 38: Costes de Personal**

### 6.1.2 Costes de Equipamiento

En este apartado se incluye la descripción del gasto realizado en el material empleado en el proyecto.

Al coste del personal hay que añadirle el coste de todas las herramientas necesarias para el proyecto. Todos estos costes no incluyen IVA.

Equipo	Coste (€)
Licencia Microsoft Project 2010	1.067,00
Licencia Microsoft Office 2010	539,00
<b>Total Actividades</b>	<b>1.606,00</b>

**Ilustración 39: Costes de equipamiento**

### 6.1.3 Coste Total

El coste total del proyecto es la suma del coste del personal más el coste del equipamiento. A esta cantidad también hay que añadirle un porcentaje de beneficio y el resultado se mostrará con IVA y sin IVA. El IVA en la actualidad supone un 21%.

- **Costes de personal. . . . . 4.755,08 euros**
- **Costes de equipamiento. . . . . 1.606,00 euros**
- **Beneficio (20%). . . . . 1.272,22 euros**
- **Subtotal. . . . . 7.633,30 euros**
- **Impuestos (21% de IVA). . . . . 1.602,99 euros**
- **TOTAL . . . . . 9.236,29 euros**

El coste total del proyecto asciende a nueve mil doscientos treinta y seis euros y veintinueve céntimos de euro.

## 6.2 Coste real

Se produce una variación entre la planificación inicial y la real de 50 horas de más, y esto repercute en el presupuesto real.

Además se han utilizado otras herramientas distintas a las iniciales pero esto no ha tenido repercusión.

### 6.2.1 Costes de Personal

En el presupuesto real se toman las mismas premisas que en el presupuesto inicial en cuanto a perfiles, valoración económica en horas, etc.

La valoración económica para estos dos perfiles, teniendo en cuenta el número de horas invertidas en la realidad es la siguiente:

	Perfil	Tarifa/ Hora	Horas Totales	Coste Total
<b>Recurso 1</b>	Ingeniero Téc. Informático Jr.	11,06	405	4.479,30
<b>Recurso 2</b>	Ingeniero Informático Sr.	22,12	40	884,80
<b>Coste Total Personal</b>			<b>445</b>	<b>5.364,10</b>

**Ilustración 40: Costes de equipamiento**

### 6.2.2 Costes de Equipamiento

Tal y como se ha comentado anteriormente las aplicaciones utilizadas finalmente han variado en relación con las iniciales pero esto no ha supuesto ningún cambio económico en este apartado.

### 6.2.3 Coste Total

Igual que pasaba en el presupuesto inicial el coste total real del proyecto es la suma del coste del personal más el coste del equipamiento. A esta cantidad también hay que añadirle un porcentaje de beneficio y el resultado se mostrará con IVA y sin IVA. El IVA en la actualidad supone un 21%.

- **Costes de personal. . . . . 5.364,10 euros**
- **Costes de equipamiento. . . . . 1.606,00 euros**

## CAPÍTULO 9

- **Beneficio (20%). . . . . 1.394,02 euros**
- **Subtotal . . . . . 8.364,12 euros**
- **Impuestos (21% de IVA). . . . . 1.756,46 euros**
- **TOTAL . . . . . 10.120,58 euros**

El coste total del proyecto asciende a diez mil ciento veinte euros y cincuenta y ocho céntimos de euro.

# Conclusiones

En este proyecto se ha descrito el proceso seguido para analizar el impacto o influencia de las noticias económicas en los valores de cierre de la Bolsa de Madrid. Para ello se ha hecho un estudio de distintas herramientas de análisis de sentimientos.

El análisis de sentimientos es una tarea muy importante hoy en día en toda red social y es muy útil, sobre todo, para conocer la opinión de la gente acerca de un tema concreto o para calificar noticias (como en este caso) en positivas o negativas y que ayuden a automatizar la previsión de ciertos comportamientos.

Una vez analizadas las noticias, se ha creado una herramienta de visualización que permite ver de una gran cantidad de información, la que al usuario realmente le interesa. Esta herramienta se ha intentado simplificar al máximo, con 3 filtros (empresa, fecha, noticia) que el usuario selecciona y el resultado es un gráfico con la variación de días anteriores y días posteriores a la noticia, el ratio de la noticia y la variación porcentual desde el día de la noticia y los días siguientes.

Este proyecto ha supuesto un gran reto para mí y me ha servido tanto para aprender mucho acerca del análisis de sentimientos y el análisis de la información como aprender a desarrollar un proyecto desde el inicio y participando en todas las fases.

He tenido que realizar un trabajo muy extenso de documentación y de conceptualización ya que es un tema en auge actualmente, con mucha información, pero con un mundo entero por descubrir. Existe mucha documentación en inglés, debido a su sencillez, pero en España aún queda mucho trabajo por hacer.

Con el desarrollo del proyecto también he adquirido muchos conocimientos acerca del ciclo de vida un proyecto. Todas las fases previas a la implementación que son tan necesarias para que el trabajo sea productivo y la labor de documentación, tan necesaria también, para reflejar todo el trabajo realizado. Otro apartado importante y que me ha aportado gran conocimiento ha sido la planificación y el presupuesto del proyecto.

Todos los conocimientos que he ido adquiriendo durante el proyecto y todas las horas invertidas buscando información han ido generando en mí una inquietud por seguir trabajando en esta línea, tanto en temas relacionados con la gestión de los datos en la red, como en la gestión de proyectos.

Espero que en un futuro, esta haya sido la base de un gran conocimiento.

# Trabajos futuros

A lo largo del proyecto han ido surgiendo varias ideas que no se han podido desarrollar en este trabajo pero se sugieren realizar en el futuro con el objetivo de aportar valor añadido a éste. Las propuestas de mejora son las siguientes:

- En este proyecto se han relacionado las noticias con las empresas cotizadas en la Bolsa de Madrid a través del nombre completo de la empresa. La primera mejora que se propone es hacer una relación que abarque más conceptos a parte del nombre de la empresa. Por ejemplo, si queremos recuperar las noticias que pueden afectar al valor de la empresa BBVA, buscaremos palabras clave como *'banca'*, *'banco'*, *'crédito'*, *'hipoteca'*, etc.
- Automatizar el proceso previo de recuperación de noticias y valores de cierre. Para ello hay que descargar las noticias y los valores diariamente en una base de datos e integrar el API de análisis de sentimiento en la herramienta para sacar la puntuación de la noticia.
- Crear un sistema de notificaciones personalizadas al usuario. Este sistema se puede parametrizar de la siguiente manera:
  - Empresas en las que el usuario tenga acciones (avisando de una previsión de caída del valor).
  - Previsiones de fuertes subidas, sugiriendo la compra de acciones.
  - Seguimiento de empresas seleccionadas por el usuario.
- Integrar otras fuentes de datos, añadiendo un peso variable dependiendo de cual sea la fuente. Por ejemplo, si integramos Twitter, no puede tener el mismo peso un tweet de una fuente oficial de otra que no lo sea.

# Glosario de acrónimos

- PLN – Procesamiento de Lenguajes Naturales
- API – Application Programming Interface
- BBVA – Banco Bilbao Vizcaia Argentaria
- IVA – Impuesto sobre el Valor Añadido
- VBA – Visual Basic for Applications
- MS – Microsoft
- BME – Bolsas y Mercados Españoles
- ID – Identificador
- ADV – Adverbio
- PP – Nombre propio
- VB – Verbo
- ART – Artículo

# Bibliografía

- HemerotecaExpansión. (s.f). Obtenido de <http://www.expansion.com/hemeroteca/2014/>
- Cierres diarios empresas cotizas de la Bolsa de Madrid en el año 2014 <http://www.bolsamadrid.es/>
- Redes Sociales, Blog Bankinter. (2014) ¿Puede Google predecir el comportamiento de las bolsas? Obtenido de <http://blog.bankinter.com/>
- Tobias Preis, Helen Susannah Moat & H. Eugene Stanley. (2013). Quantifying Trading Behavior in Financial Markets Using Google Trends. Obtenido de <http://www.nature.com/srep/2013/130425/srep01684/full/srep01684.html>
- Fernando Berzal. El ciclo de vida de un sistema informático. Obtenido de <http://elvex.ugr.es/idbis/db/docs/lifecycle.pdf>
- BBVA. (s.f.). ActiBVA. Obtenido de <http://www.actibva.com/>
- Aprendizaje profundo [https://es.wikipedia.org/wiki/Aprendizaje\\_profundo](https://es.wikipedia.org/wiki/Aprendizaje_profundo)
- Información de la aplicación Semantria <https://en.wikipedia.org/wiki/Semantria>
- Mark Cieliebak, Oliver Dür, and Fatih Uzdilli. Zurich University of Applied Sciences Winterthur, Switzerland. Potential and Limitations of Commercial Sentiment Detection Tools <http://ceur-ws.org/Vol-1096/paper4.pdf>
- Información sobre la metodología ágil SCRUM y datos de interés <https://es.wikipedia.org/wiki/Scrum>
- Definición Pruebas Funcionales [https://es.wikipedia.org/wiki/Pruebas\\_funcionales](https://es.wikipedia.org/wiki/Pruebas_funcionales)



# Anexo I. Cálculo Salario Profesionales

A través de ActiBVA (BBVA), que es una comunidad impulsada por BBVA y desarrollada por Weblogs SL cuyo objetivo es acercar la información financiera a todos los usuarios, se ha realizado el cálculo del salario neto anual de cada profesional asignado para comprobar su viabilidad en el proyecto.

Los cálculos son los siguientes:

- Ingeniero Técnico Informático Jr.

**ActiBva Sueldo n...**

**Entrada de datos**

**Datos generales:**

- Estado civil: Soltero
- Retribución dineraria: 20.000 €
- Retribución en especie: 0 €
- ¿Se deduce por vivienda habitada?: No

**Datos para calcular deducción:**

**Año de nacimiento:** 1987

**Descendientes:**

- Menores de 3 años: 0
- Menores de 25 años: 0

**Ascendientes:**

- Mayor de 65 años: 0
- Mayor de 75 años: 0

**Seguridad Social**

**Categoría profesional:** Ingenieros técnico, peritos y ayudantes t...

**Resultados**

**Retribución bruta anual:** 20.000,00 €

**Retención IRPF anual:** 2.456,00 €

**Seguridad Social anual:** 1.270,00 €

**Sueldo neto:**

- 12 pagas: ☒ Mensual: 1.356,17 €
- 14 pagas: ☐
- 16 pagas: ☐
- Paga extra: 0,00 €

© 2015 Afinet (Afi). Simulación ajustada a la normativa vigente a 31/01/2015.

**Ilustración 41: Cálculo salario Ingeniero Técnico Informático Jr.**

- Ingeniero Informático Sr.

**Activa Sueldo n...**

**Entrada de datos**

**Datos generales:**

- Estado civil: Soltero
- Retribución dineraria: 40.000 €
- Retribución en especie: 0 €
- ¿Se deduce por vivienda habitada?: No

**Seguridad Social**

- Categoría profesional: Ingenieros técnico, peritos y ayudantes t

**Datos para calcular deducción:**

Año de nacimiento: 1987

**Descendientes**

- Menores de 3 años: 0
- Menores de 25 años: 0

**Ascendientes**

- Mayor de 65 años: 0
- Mayor de 75 años: 0

**Resultados**

- Retribución bruta anual: 40.000,00 €
- Retención IRPF anual: 8.176,00 €
- Seguridad Social anual: 2.540,00 €
- Sueldo neto:**
  - 12 pagas: ☒ Mensual: 2.440,33 €
  - 14 pagas: ☐
  - 16 pagas: ☐ Paga extra: 0,00 €

© 2015 Afinet (AFI). Simulación ajustada a la normativa vigente a 31/01/2015.

**Ilustración 42: Cálculo salario Ingeniero Informático Sr.**

El sueldo que nos muestra este cálculo es mensual, el resultado que se ha obtenido parece óptimo por lo que a continuación se va a estimar en horas.

Para este cálculo se utiliza un estándar de horas anuales estipulado en 1.808. Este número de horas se debe a que un año tiene 365 días, de los que se trabajan 226 aproximadamente (22 días de vacaciones, 52 fines de semana (104 días), 13 días festivos aprox.) y el número de horas de trabajo diarias son 8, por lo tanto:

$$(365 - 139 \text{ (vacaciones, fines de semana y festivos)}) * 8 = 1.808$$

Teniendo en cuenta esto, la tarifa es la siguiente:

- Ingeniero Técnico Jr.:  $20.000 / 1.808 = 11,06 \text{ €/h.}$
- Ingeniero Sr.:  $40.000 / 1.808 = 22,12 \text{ €/h.}$